Parametric Potential-Outcome Survival Models for Causal Inference

---

A thesis submitted in partial fulfilment of the
requirements for the Degree of
Doctor of Philosophy
by
Zhaojing Gong

University of Canterbury
October 2008

To my parents

Abstract

Estimating causal effects in clinical trials is often complicated by treatment noncompliance and missing outcomes. In time-to-event studies, estimation is further complicated by censoring. Censoring is a type of missing outcome, the mechanism of which may be non-ignorable. While new estimates have recently been proposed to account for noncompliance and missing outcomes, few studies have specifically considered time-to-event outcomes, where even the intention-to-treat (ITT) estimator is potentially biased for estimating causal effects of assigned treatment. In this thesis, we develop a series of parametric potential-outcome (PPO) survival models, for the analysis of randomised controlled trials (RCT) with time-to-event outcomes and non-compliance. Both ignorable and non-ignorable censoring mechanisms are considered. We approach model-fitting from a likelihood-based perspective, using the EM algorithm to locate maximum likelihood estimators. We are not aware of any previous work that addresses these complications jointly. In addition, we give new formulations for the average causal effect (ACE) and the complier average causal effect (CACE) to suit survival analysis. To illustrate the likelihood-based method proposed in this thesis, the HIP breast cancer trial data (Baker, 1998; Shapiro et al., 1988) were re-analysed using specific PPO-survival models, the Weibull and log-normal based PPO-survival models, which assume that the failure time and censored time distributions both follow Weibull or log-normal distributions. Furthermore, an extended PPO-survival model is also derived in this thesis, which permits investigation into the impact of causal effect after accommodating certain pre-treatment covariates. This is an important contribution to the potential outcomes, survival and RCT literature. For comparison, the Frangakis-Rubin (F-R) model (Frangakis & Rubin, 1999) is also applied to the HIP breast cancer trial data. To date, the F-R model has not yet been applied to any time-to-event data in the literature.

Acknowledgments

Many people have given me their generous and kind assistance during these last few years. Without their support, this thesis could not have been realised. I would like to thank them here for the roles they played, and the impact they have made on this research, and on me.

Firstly, I would like to thank my primary supervisors: Associate Professor Irene Hudson of the School of Mathematics and Statistics at the University of South Australia and Dr. Patrick Graham of the Department of Public Health and General Practice, University of Otago, Christchurch. I thank them for their direction and their continuous belief in me, always pushing me further than I ever thought I could achieve. Without their patient guidance, I would not have been able to embark upon this journey, let alone reach the completion of this thesis. Particularly, my many thanks to Associate Professor Irene Hudson who helped proofread my thesis, albeit from across the Tasman, with many drafts of written corrections and advisements by email scan. And to Dr. Patrick Graham who inspired and guided me into such an interesting area of study as causal inference. Thanks to him for providing me with many helpful suggestions to overcome obstacles I encountered during creation of these models.

I would also like to thank Dr James O'Malley, of the Department of Health Care Policy, Harvard Medical School, USA, for his generous assistance and helpful suggestions, and Dr. Chris Hann of the Department of Mechanical Engineering, University of Canterbury, for his useful tips on how to use MATLAB commands more effectively. I want to thank Professor David Wall and Professor Mike Steel for taking on the role of my Administrative supervisor successively, during the last two years after my senior supervisor, Irene Hudson, transferred to the University of South Australia. Especially, thanks to Professor Mike Steel for his helpful comments on my thesis.

I would also like to thank Dr. Stuart G. Baker, National Cancer Institute of USA, for kindly organising the HIP data for this research. I thank also

Diane Erwin, the assistant of Dr. Baker, for answering all the questions I asked about the HIP data with patience.

Thanks to Paul Brouwers and Steve Gourdie, the two brilliant IT technicians of the Maths and Stats Department, for their expert guidance in the many programs and software issues I've encountered, even though some of them were beyond their direct responsibility.

I want to thank Susan Bidwell, my colleague in the Department of Public Health and General Practice, University of Otago, Christchurch, for kindly volunteering to proofread my thesis. For your generosity, I am forever grateful. My thanks to Megan Foster for proofreading my thesis also. And to In Kang, my office mate whom I have shared good times with during the years.

I would like to thank my sponsors: New Zealand Institute of Mathematics and its Application (NZIMA) for my three year Ph.D. scholarship, and the New Zealand Statistics Association (NZSA) and the Department of Mathematics and Statistics for also sponsoring me to attend relevant conferences.

To my parents, for their endless support to me. To my sisters, Qing and Xing, for their companionship and motivation during the early years. To my daughter, Yanxin Zhu, for being so understanding and independent.

Most importantly, I would like to thank my husband, Fanghua Zhu, for his love and support right from the beginning, for bearing with me through the tough times and motivating me to do better. Your continuous support for me has been critical to the accomplishment of this thesis, and the enrichment of my life. Thank you for everything.

Finally, thanks to God for maintaining and providing me with such wonderful people along this journey. It has been a wonderful journey indeed.

Publications Related to this Study

Paper in preparation

Gong, Z.J.; Hudson, I.L.; Graham, P.; Hann, C. and O'Malley, J.
Likelihood Methods for Randomized Experiments with All-or-None Compliance for Time-to-Event Data (pp 00-85)
In prep for the *Biostatistics* Journal.

Refereed conference proceedings

Gong, Z.J.; Hudson, I.L. and Graham, P. (2005) **A Bayesian Causal Cox Model (CCM) for All-or-Nothing Compliance** In proceedings of the *20th International Workshop in Statistical Modelling, Sydney, Australia.* Francis, A.R.; Matawie, K.M.; Oshlack, A. and Smyth, G.K.(Eds), 227-230.

Conference presentations

Gong, Z.J.; Hudson, I.L. and Graham, P. (2005)
A Bayesian Causal Accelerated Failure Time Model
New Zealand Statistical Association (NZSA) 2005 Conference, 4-6 July, Otago University, Dunedin.

Gong, Z.J.; Hudson, I. ; Graham, P. ; Hann, C. and O'Malley, J. (2006)
A Likelihood Method for Randomized Time-to-Event Studies with All-or-None Compliance: A Re-analysis of the HIP Study
Australia Statistical Conference (ASC)/New Zealand Statistical Association (NZSA) Conference 2006, 3-6 July, SKYCITY, Auckland.

List of abbreviations

ACE Average causal effect;

AFT Accelerated failure time model;

AIC Akaike information criterion;

CACE Complier average causal effect;

EM Expectation and maximisation algorithm;

F-R Frangakis-Rubin model;

HIP Health insurance plan breast cancer trial;

IGN-L Ignorable censoring mechanism parametric potential-outcome survival log-normal model;

IGN-W Ignorable censoring mechanism parametric potential-outcome survival Weibull model;

ITT Intention-to-treat;

IV Instrument variable;

MAR Missing at random;

MCAR Missing completely at random;

NIGN-LL Non-ignorable censoring mechanism parametric potential-outcome survival log-normal model;

NIGN-WW Non-ignorable censoring mechanism parametric potential-outcome survival Weibull model;

NMAR Not missing at random;

OIM Observed information matrix;

PH Proportional hazard model;

POM Potential-outcome model;

PPO-survival model Parametric potential-outcome survival model;

RCM Rubin causal model;

RCT Randomised controlled trial;

SUTVA Stable unit treatment value assumption.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 A case-study in medical research: HIP breast cancer study

In medical research, investigators often need to determine whether a medical treatment or a detection procedure is effective. The following questions often pertain: "What is the efficacy of a certain drug in a given population?" or "What fraction of deaths from a given disease could have been avoided by a given treatment or policy or intervention?"

For convenience, we start from a real case-study of a randomised breast cancer clinical trial, known as the Health Insurance Plan (HIP) trial, which was undertaken in Greater New York (Shapiro, 1977; Shapiro et al., 1982, 1988; Baker, 1998).

1.1.1 The HIP breast cancer study

In the early 1960s, the mammography technique was developed to detect breast cancer through screening procedures. At the time of its release, it was not certain if screening indeed reduced the mortality rate (i.e. death from breast cancer). To investigate this question, the Health Insurance Plan (HIP) of Greater New York, USA, conducted a Screening Examination Trial commencing in 1963 (Shapiro, 1977; Shapiro et al., 1982, 1988; Baker, 1998). In the HIP breast cancer trial, participants were women who had enrolled as members of the Health and Insurance Plan (HIP) of Greater New York for one or more years[1]. The trial continued from 1963 to 1986. All participants were over the age of 40 at entry into the trial and had no breast cancer symptoms when they enrolled. All participants in the HIP trial were followed up in the

[1]From the date the programme started in 1963.

long term, up to 18 years from their date of entry.

In the HIP trial, a total of 60,696 women were eligible. Of these,30,131 were randomly assigned to the screening group and 30,565 were assigned to the control group. Women in the *screening* group were offered free annual breast examinations four times. These examinations consisted of film mammography (cephalocaudal and later views of each breast), a clinical examination of the breast by a physician (usually a surgeon), an interview for demographic and other background information; and a health history, while the women in the control group only received their usual health care.

For simplicity, we introduce some notation (Table 1.1) to illustrate the HIP trial. Let N denote the sample size and N_z (for $z = 1$ and 0) denote the number of individuals in the relevant treatment group. Here $Z_i = z$ is the randomised assigned treatment (1 for the screening group, and 0 for the control group). We then have $N = 60,696$; while $N_1 = 30,131$ and $N_0 = 30,565$ for the HIP trial.

For the 9,984 women in the screening group who refused to take the screening examination, it is considered that these women switched off their assigned treatment directly. Let $D_i(Z_i)$ denote the actual received treatment when the individual is assigned to the treatment group $Z_i = z$ (for $z = 1$ and 0); we can then write $Z_i = 1$ and $D_i(1) = 0$ for these women. This implies that the HIP trial falls into the so-called randomised controlled trial (RCT) with all-or-none compliance (Section 1.2) scenario (Baker, 1997).

Table 1.1: Summary statistics: HIP data.

	Total	Screen invited $Z = 1$	Control $Z = 0$	Screen received $Z = 1, D = 1$	Screen refused $Z = 1, D = 0$	No Screen received $D = 0$
Sample size	60,696	30,131	30,565	20,147	9,984	40,549
BC patients	2,325	1,171	1,154	835	336	1490
Deaths of BC	786	372	414	247	125	539
Mean of observed time $\bar{Y}$	14.051	14.1578	13.9457	14.905	12.6499	13.6266
Proportion of $\bar{R} = 1$	0.0129	0.0123	0.0135	0.0123	0.0125	0.0133
Proportion of $\bar{Z} = 1$	0.4964	1	0	1	1	0.2462
Proportion of $\bar{D} = 1$	0.3319	0.6686	0	1	0	0

$R = 1$: event occurs;
$Z = 1$: randomly assigned to the screening group;
$D = 1$: received the screening examination.

During the HIP trial observation period (1963-1986), a total of 2,325 women were diagnosed with breast cancer (BC) (1171 in the screening group and 1154 in the control group); the 786 who died due to breast cancer are denoted by ($R_i = 1$). Of these, 372 were in the screening group and 474 were in the control group. In addition, in the HIP trial, the date of enrolment for all women was recorded, as well as their date of death (if an individual died from breast cancer) or their date of the last follow-up (if an individual did not die from breast cancer). The length of time, from the date of enrolment till the date of death from breast cancer (if an individual died from breast cancer) was then observed and known as the so-called event time or failure time, denoted by T_i. In general, when the outcome of interest is event time, then the data are called *time-to-event data* and the pertinent analysis is termed *time-to-event analysis* (Section 1.3).

Note that individuals who did not die from breast cancer had dates of death from breast cancer as unknown. Therefore their event time or failure times T_i were not available in the HIP trial. This situation is termed *censoring* (Section 1.3), and these individuals are known as the censored individuals. For censored individuals, we note that, instead of their failure time, it is their censoring times, denoted by C_i, that are observed, namely the length of time from the date of enrolment till the date of last follow-up. Hence, no matter whether an individual in the HIP trial died from breast cancer or not, at least one of either the failure time or censoring time is observed. This is the so-called observed time $Y_i(z)$ and it can be written as $Y_i(z) = \min(T_i(z), C_i(z))$, where z denotes the assigned treatment, that is $Z_i = z$.

In the HIP trial, the mean of the observed time, $\bar{Y} = \sum_i^N Y_i(z)/N$ (for all individuals, including both $z = 1$ and $z = 0$), is $\bar{Y} = 14.051$; whereas for the screening group ($Z_i = 1$) (or the control group ($Z_i = 0$)), the mean values of observed time are $\bar{Y}(1) = \sum_{i\in\{i:Z_i=1\}}^N Y_i(1)/N_1 = 14.1578$ (or $\bar{Y}(0) = \sum_{i\in\{i:Z_i=0\}}^N Y_i(0)/N_0 = 13.9457$). The mortality of breast cancer in the screening group was $0.0123/18 = 6.83E-4$ (per year); whereas it was $0.0135/18 = 7.5E-4$ (per year) for the control group. Basic statistics on the HIP data are given in Table 1.1. For convenience, all notation used in this thesis is given in Table 1.2.

Table 1.2: Primary notation.

		Potential and Observable data
Term		Data for subject i
Assigned treatment		$Z_i = z$ ($1 =$ intervention, $0 =$ control)
Compliance type		$U_i = u$ ($c =$ complier,$n =$ never-taker, and $a =$ always-taker)
	Potential data	Observable Data
Received treatment	$\mathbf{D}_i = [D_i(1),\ D_i(0)]$	$D_i^{obs} = zD_i(1) + (1-z)D_i(0)$
Failure time	$\mathbf{T}_i = [T_i(1),\ T_i(0)]$	$T_i^{obs} = zT_i(1) + (1-z)T_i(0)$
Censored time	$\mathbf{C}_i = [C_i(1),\ C_i(0)]$	$C_i^{obs} = zC_i(1) + (1-z)C_i(0)$
Censored indicator	$\mathbf{R}_i = [R_i(1),\ R_i(0)]$	$R_i^{obs} = zR_i(1) + (1-z)R_i(0)$
Observed time	$Y_i(z) = min(T_i(z),\ C_i(z))$	$Y_i = zY_i(1) + (1-z)Y_i(0)$
		Population parameters
Parameter	Notation	Subpopulation
Model parameter	$\beta = (\theta_{\mathbf{T},\mathbf{uz}}, \theta_{\mathbf{C},\mathbf{uz}}, \pi_{u,zd})$	$(U,Z)_i = (u,z)$
Survival probability	$S_{uz}(t)$	$(U,Z)_i = (u,z)$
Proportion $U_i = u$	π_u	All
Proportion $(Z,D)_i = (z,d)$	λ_{zd}	All
		Data summaries
Statistics	Notation	Subsample
Sample size	n_{zdr}	$(Z,D,R)_i = (z,d,r)$
Mean observed time	$\bar{Y}_{zd}$	$(Z,D)_i = (z,d)$
Proportion of event	$\bar{R}$	$(Z,D,R)_i = (z,d,r)$

The main task of the HIP trial was to determine to what degree mammography was effective in reducing breast cancer mortality. However, this question is not easily answered. This thesis aims to find an appropriate way to answer this question rigorously, using the concepts of *non-compliance* and of *all-or-none compliance*, defined in Section 1.2 via established literature. Section 1.3 addresses a special type of data: the so-called time-to-event data and their unique features. *Causal inference* is introduced briefly in Section 1.4; some common methods used in causal inference for noncompliance analysis are addressed in Sections 1.6 and 1.7.1. The most important covariate in noncompliance analysis, the so-called *compliance type* U_i, is defined in Section 1.7.2. The aim of this thesis is outlined in Section 1.8, and an overview of the thesis is given in Section 1.9.

1.2 All-or-none compliance

About one-third of participants in the screening group of the HIP trial refused the examination, which indicates that not all participants in the treatment group had taken up the offer of a screening examination. This phenomenon is termed *treatment switching* from an assigned treatment, that is, participants involved in a trial do not follow their assigned treatment properly. This is also referred to as *noncompliance* in the literature (Dunn & Goetghebeur, 2005a; White, 2005; Dunn et al., 2005b; Baker & Kramer, 2005).

When *treatment switching* occurs immediately after randomisation, this is known as *all-or-none compliance*, a special type of *noncompliance*, in which it is supposed that every patient either does or does not comply with their assigned treatment completely, and no partial compliance exists. The HIP breast cancer trial falls into the *all-or-none compliance* category (Baker, 1997). *All-or-none compliance* studies have received much attention over the last decade (Imbens & Rubin, 1997; Baker, 1998; Frangakis & Rubin, 1999; Hirano & Imbens, 2000; Jo, 2002a,b; Loeys & Goetghebeur, 2003; Levy et al., 2004; Peng et al., 2004; O'Malley & Normand, 2005; White, 2005; Gong et al., 2005).

1.3 Time-to-event data

In medical research, the outcome of interest is often related to time; such as the time of death due to a given disease, the time to relapse, or the time to a specific symptom appearing. Time-to-event data refers to the case where the outcome of interest is the time at which individuals experienced a pre-specified event (such as death, relapse or failure status).

As noted earlier, in the HIP breast cancer trial, the pre-specified event is death from breast cancer, and the outcome of interest is the time when an individual dies from breast cancer, which is the event time (or failure time). This is also called survival time, because the event time in the HIP trial implies that a given individual has survived until the event (death) occurs (given breast cancer death eventuates).

When an individual's failure time cannot be observed because the individual did not experience the event of interest (during the observation time), we say her/his failure time is censored, and call such individuals censored individuals. In time-to-event studies, it would not be realistic to expect that all individuals involved in the trial have an available failure time, since it is likely that some may not experience the pre-specified event in the given observation time frame or follow-up time. Hence, *censoring* is the primary concept that has to be accommodated for in time-to-event studies. Since, for censored individuals, their real failure time would not be observed, this can be treated as a non-response. On the other hand, their censoring time is observed. In this case, we still can gather information on them via their observed censoring time. For example, if an individual in the HIP trial did not die from breast cancer by the date that follow-up ended, their actual survival time cannot be observed. However, we know that this individual has a survival time at least after that day. Let $C_i = t$ denote the censoring time, i.e. the last follow-up recorded time. Then this individual's survival time T would be greater than t, which implies that $T_i > t$ holds. Because of this feature, time-to-event data is at least partially observable.

1.4 Causal inference

1.4.1 Causal inference in general

Causal inference, roughly speaking, seeks to infer the causation of interest through observations (Pearl, 2000; Graham, 2000; Geng & Jing, 2002). For example, Bradford Hill's criteria (Graham, 2000) is a well-known approach of causal inference in epidemiology. In this approach, Hill gave nine causal criteria for judging if there is evidence for causation. The nine causal criteria are: *strength of association, consistency, specificity of association, temporality, biological gradient, plausibility, coherence, experiment and analogy.* These criteria typically determine whether the associations between the exposures and outcomes represent a causal effect (as judged by investigators) (Graham, 2000). *Causal inference* generally focuses on learning the causal relationships from raw data, whereby *statistical inference* is often employed (Pearl, 2000). *Statistical inference* is usually concerned with associational inference (Holland, 1986), as its purpose is mainly to discover how the response variables (outcomes) are associated with other covariate variables (exposures) for each subject in a population. As in Hill's criteria-based approach, statistical methods are used only for finding associations between exposures and outcomes (statistical inference), and are not used for inferring causation from observations (causal inference). Therefore, *statistical inference* is not the same as *causal inference*, even though there is certainly a parallel in their inferential process.

Causal inference involves the extraction of information about comparisons, such as decisions in economic, educational and medical research, and also in public health or social policy. This is dependent on the appropriate evaluation of competing treatments or policies (Holland, 1986; Frangakis & Rubin, 2002).

Statistical studies of causal inference have progressed rapidly during the last two decades (details given later in this section). An increasing number of researchers accept the idea that causality should play a significant role in scientific work (Pearl, 2000) and the growing emphasis, over recent time, on evidence-based practice (Loeys & Goetghebeur, 2003; Mealli et al., 2004; Baker, 1998) has led to an increased commitment to studies of clinical

effectiveness on the part of clinicians, researchers and policy makers. It is noteworthy that the rapid development of computational techniques makes complicated computations involved in causal inference more tractable (Pearl, 2000).

Causal inference is a large topic and encompasses many methodologies applied to many applications. This thesis focuses on a specific methodology of causal inference which is based on the potential outcomes framework (Dawid, 2000; Hirano & Imbens, 2000; Hogan & Daniels, 2002; Barnard et al., 2003; Greenland, 2004).

1.4.2 Potential outcomes framework

A milestone in causal inference is the emergence of the potential outcomes framework, in which all possible outcomes, including those that are observable and unobservable (counterfactuals), are considered simultaneously. Causal inference has a history that goes back to the early work of Neyman in the 1920s. He first used the potential outcomes framework in his Ph.D. thesis in 1923 (Rubin, 2005). However the potential outcomes framework had not attracted much attention in statistical area until Rubin's work (Rubin, 1974, 1978). During the last few decades, more and more researchers have considered the potential outcomes framework, including frequentists (Frangakis & Rubin, 1999; Little & Rubin, 2000; Joffe, 2001; Joffe et al., 2003; Loeys & Goetghebeur, 2003; Barnard et al., 2003; Peng et al., 2004a; O'Malley & Normand, 2005) and Bayesians (Imbens & Rubin, 1997; Graham, 2000; Hirano & Imbens, 2000; Hamilton, 2001; Chib & Hamilton, 2002; Hogan & Daniels, 2002; Gong et al., 2005).

In brief, the potential outcomes framework can be described as a framework in which the outcome of interest appears as a vector, which is the so-called *potential outcomes vector* (Imbens & Rubin, 1997; Greenland, 2004).

Potential outcomes vector

The *potential outcomes vector*, as its name suggests, is a vector corresponding to all possible responses from all exposure or treatment levels, regardless of whether the individual is exposed or not in practice. This concept can be

described precisely via mathematical notation, as follows.

Suppose there are p potential exposure or treatment levels in a RCT. In general, we use $\mathbf{Y}_i$ to denote the potential outcomes vector for the *ith* individual, where $\mathbf{Y}_i = [Y_i(0),\ Y_i(1),\ \ldots,\ Y_i(p-1)]$. For convenience and without loss of generality, let us address the simplest case with $p = 2$ (when $p > 2$, it is straightforward to extend the arguments we used from the simplest case). Then the potential outcomes vector $\mathbf{Y}_i$ has two components $[Y_i(0),\ Y_i(1)]$, where $Y_i(0)$ denotes the outcome of the *ith* individual if assigned to the control group and $Y_i(1)$ represents the outcome of the *ith* individual if assigned to the treatment group. In our study, the possible treatments (in a RCT) are only the so-called treatment and control. Note that for each individual, only one of the outcome components of $Y_i(0)$ and $Y_i(1)$ is observed.

1.4.3 Causal inference in the potential outcomes framework

During the last decade, a methodology based on the potential outcomes framework (Dawid, 2000; Hirano & Imbens, 2000; Hogan & Daniels, 2002; Barnard et al., 2003; Greenland, 2004) in causal inference attracted much more attention (Section 1.4.2). Apart from its application to epidemiology and medical research (Shrout, 1998; Schuck & Widom, 2001; Waitzkin et al., 2001; Holman et al., 2001; Iribarren et al., 2001; Kaufman & Cooper, 2001; Schochat & Seidel, 2001; Aiello & Larson, 2002; Little & Rubin, 2002; Suter et al., 2002; Weed, 2002; Acquavella et al., 2003; Breslow, 2003; Joffe & Brensinger, 2003b; Li & Sung, 2003; Terracini & Zanetti, 2003; Hernan et al., 2005), this methodology has also been applied to economics (Dehejia & Wahba, 2002; Persson & Tabellini, 2002), education (Barnard et al., 2003; Jo, 2002a), psychology (White, 2001; Cobos et al., 2002; Jo, 2002; Hassin et al., 2002); and social science (Draper, 1995; Yau & Little, 2001).

In brief, this methodology regards the outcome of interest as a vector that corresponds to each potential treatment level (Greenland, 2004) rather than a scalar, and defines the *causal effect* as a quantity which contrasts the components of the potential outcomes vector (Rubin, 1978; Holland, 1986; Graham, 2000), thereby describing the effect caused by a certain intervention. Its formal definition is given in Chapter 2.

*Causal effect*s may vary in different applications: for example, the effect of education on employment and earnings, the effect of an employment training programme on subsequent labour market histories, and the effect of a job training intervention for preventing deterioration in an individual's mental health as a result of job loss. In the field of medicine and health sciences, the usual effect of interest is that of a given drug or therapy for a specific disease, such as the effect of the drug 6-mercaptopurine (6-MP) for children with acute leukaemia (Klein, 1997) or the effect of treatment via resection and chemotherapy for liver metastases due to colorectal cancer (Loeys & Goetghebeur, 2003). In the HIP breast cancer trial, the *causal effect* of interest is the effect of the breast screening examination in reducing breast cancer mortality.

1.4.4 Fundamental problem of causal inference

In practice, however, with respect to the potential outcomes vector $\mathbf{Y}_i$, only one of its components is observable. This is because an individual only receives one possible treatment in a RCT. This problem is defined as the fundamental problem of causal inference (Holland, 1986). Consequently, the *causal effect* is defined as a contrast of two potential outcomes for each individual. It is also unobservable. To deal with this problem, the most plausible approach proposed to date is to use randomisation (Holland, 1986; Imbens & Rubin, 1997).

1.5 Randomisation

The *randomisation* technique was introduced by Sir RA Fisher in the 1920s, and this procedure creates a fair balance between the randomly assigned groups, including those with unknown confounders.

The key point of randomisation is that no-one can determine which treatment group an individual should be in unless we obtain a randomised assignment (Z_i) by a randomly drawn given distribution (such as the bernoulli, binomial or uniform distributions). This is similar to the procedure of tossing a coin at only two treatment levels, e.g. treatment ($Z_i = 1$) and control ($Z_i = 0$). The important role of randomisation is to yield comparable treat-

ment groups without bias. Therefore the treatment and control groups in a RCT can be viewed as drawn randomly from the same population. This property of randomisation is useful in solving the fundamental problem of causal inference. Further discussion is given in Chapter 2.

Despite randomisation being a useful tool for solving the fundamental problem of causal inference, it requires that no noncompliance exists in RCTs. One simplest solution is to ignore all noncompliance issues in a RCT, though this is not a highly plausible approach.

1.6 Intention-to-treat (ITT) analysis

In the early causal inference literature, the issue of imperfect compliance (noncompliance) was ignored. This means that all individuals involved in a randomised controlled trial (RCT) were supposed to be intent on following their randomised assignment treatment properly. The noncompliance issues were not considered. This scenario is the well known intention-to-treat (ITT) scenario (Angrist et al., 1996; Imbens & Rubin, 1997; Frangakis & Rubin, 1999; Heitjan, 1999; White, 2005).

White (2005) described the typical *intention-to-treat* (ITT) analysis as the situation that compares outcomes between the groups as randomized, ignoring the actual treatment(s) received.

Sommer & Zeger (1991) defined *biologic efficacy* as the effect of treatment for all individuals who receive the treatment to which they were assigned. This implies that all individuals in the RCT comply well with their randomised assigned treatment. Biologic efficacy then measures the biologic action of treatment among so-called compliant individuals or compliers (Section 1.7.2).

Intention-to-treat comparisons estimate the *programmatic effectiveness* or short *effectiveness* of a treatment rather than its *biologic efficacy* (Sommer & Zeger, 1991; Loeys & Goetghebeur, 2003). This is true especially in imperfect RCTs, where some participants do not properly comply with their randomised assigned treatment .

Only in perfect RCTs, wherein all participants ideally follow their randomised assignment exactly and where no treatment switch occurs, does the

intention-to-treat (ITT) approach provide an estimator of *biologic efficacy* and *effectiveness* of treatment without bias (Loeys & Goetghebeur, 2003; O'Malley & Normand, 2005; Frangakis & Rubin, 1999). Otherwise, ITT analysis may provide biased estimators of *biologic efficacy* in the analysis of data from non-perfect RCTs. This is because the ITT approach ignores the noncompliance issue and estimates the *causal effect* of assigned treatment rather than the *causal effect* of actually received treatment. This is inherently a mixture of the effect of the received treatment and of the absence of the treatment (Loeys & Goetghebeur, 2003; Frangakis & Rubin, 2002).

1.7 Noncompliance studies

1.7.1 Instrumental variables

A RCT where there is a high proportion of individuals who do not comply with their assigned treatment is likely to be viewed as a RCT with non-ignorable noncompliance. Such is the case in the HIP breast cancer trial, where up to one third of participants were non-compliant (Baker, 1998; Shapiro et al., 1988). For these type of RCTs, the ITT approach may provide a biased estimator of the effect of treatment.

In the literature for noncompliance (Angrist et al., 1996; Imbens & Rubin, 1997; Frangakis & Rubin, 1999; Hogan & Daniels, 2002; O'Malley & Normand, 2005; Gong et al., 2005), there is a so-called instrumental variable (IV) approach, which is widely used in economics and was first introduced to statisticians by Angrist et al. (1996). This IV approach has been accepted by many statisticians. Instrumental variables (IVs) are such variables that are only included in some equations (as they are explicitly excluded from others), and are correlated with some outcomes only through their effect on these subset of variables (Angrist et al., 1996).

In brief, the assigned treatment, $Z_i = z$ ($z = 1$ for the treated group and $z = 0$ for the control group), can be viewed as an *instrumental variable* (IV); $\mathbf{D}_i = [D_i(1), D_i(0)]$ is used to denote the potential received treatment, where $D_i(1) = 1$ (or $D_i(1) = 0$) indicates that an individual i assigned to the treated group $Z_i = 1$ would (or would not) actually receive the treatment. In contrast, $D_i(0) = 1$ (or $D_i(0) = 0$) represents that an individual i, assigned to

the control group $Z_i = 0$ would (or would not) actually receive the treatment. Hence in a two-arm RCT, only $D_i(1)$ or $D_i(0)$, one of the components of $\mathbf{D}_i$ denote by $D_i^{obs} = D_i(Z_i)$ is observable. The unobservable component of $\mathbf{D}_i$ can then be written as $D_i^{miss} = D_i(1 - Z_i)$ because Z_i is a binary. The miss symbol indicates that when an individual is assigned to the treated group ($Z_i = 1$), we would never be able to know his/her actual received treatment which corresponds to the control group ($1 - Z_i = 0$). Here $D_i(1 - Z_i)$ is a counterfactual variable and can never be observed in practice (Imbens & Rubin, 1997).

In the case of a perfect compliance trial, the received treatment, $D_i(z)$, for each individual would equal $Z_i = z$; either 1 or 0 in the two arms of a RCT. For example, if the HIP breast cancer trial were a perfect compliance trial, there would be no participants who were assigned to the screening group and who then refused the preferred screening examination. This is equivalent to the condition that each participant would take the screening examination ($D_i(1) = 1$) if she was assigned to the screening group ($Z_i = 1$), or would just receive normal health care ($D_i(0) = 0$) if she was assigned to the control group ($Z_i = 0$). In this case, there is little advantage of the instrumental variable (IV) approach, since $D_i(z) = z$ (where $Z_i = z$ and for $z = 1$, or 0.).

In the imperfect compliance trial, $D_i(z)$ can differ from $Z_i = z$ for various reasons. These may include cases where some participants refuse to receive the offered treatment, e.g. the screening examination in the HIP trial, (Shapiro et al., 1988; Baker, 1998); or, for example, in the ECOG trial E9288 (Loeys & Goetghebeur, 2003), where some patients with liver metastases (from colorectal cancer), assigned to the treatment group (implantation of an arterial device) had not been implanted (in the ECOG trial E9288) for reasons related to survival . In these kinds of RCTs, the assigned treatment Z_i, does not indicate whether the individual receives the treatment or not, but $D_i(z)$ does. In other words, the assigned treatment $Z_i = z$ only works through $D_i(z)$; hence Z_i is regarded as an instrumental variable (IV) (Angrist et al., 1996).

Similarly, the potential outcomes vector $\mathbf{Y}_i$ can also be expressed via the instrumental variable (IV) ($Z_i = z$); thereby $\mathbf{Y}_i = [Y(1, D_i(1)), Y(0, D_i(0))]$. It is not difficult to understand that the effect of assigned treatment, $Z_i = z$

on $\mathbf{Y}_i$, happens only through the effect of the treatment received $D_i(z)$ on $\mathbf{Y}_i$, (Imbens & Rubin, 1997; O'Malley & Normand, 2005; Graham, 2000; Gong et al., 2005) since the outcome $Y_i(z, D_i(z))$ can only be affected by $D_i(z)$, which is the actual received treatment. This implies that for an individual, no matter what the value of her assigned treatment $Z_i = z$ (for $z = 1$, or 0) is, as long as her observed and unobserved received treatments have the same value (i.e. $D_i(1) = D_i(0)$) then the components of her potential outcomes vector should be the same. That is:

$$Y_i(1, D_i(1)) = Y_i(0, D_i(0)) \quad \text{if } D_i(1) = D_i(0). \tag{1.7.1}$$

Equation (1.7.1) is equivalent to the assumption of *exclusion estriction* (Angrist et al., 1996; Imbens & Rubin, 1997; Frangakis & Rubin, 1999), its definition is given in Section 5.2. Since in practice only one of $Y_i(1, D_i(1))$ and $Y_i(0, D_i(0))$ is observable, we can then write it as $Y_i^{obs} = Y_i(Z_i, D_i^{obs}) = Y_i(Z_i, D_i(Z_i))$. The unobservable component of the potential outcomes vector $\mathbf{Y}_i$ can thus be treated as missing and we write it as $Y_i^{mis} = Y_i(1-Z_i, D_i^{mis}) = Y_i(1 - Z_i, D_i(1 - Z_i))$.

1.7.2 Compliance types

For noncompliance analysis, individuals' compliance behaviour in RCTs is a very important issue. In the literature, a variable defined to describe such behaviour is commonly known as the *compliance type* (Imbens & Rubin, 1997), *compliance state* (O'Malley & Normand, 2005), or *principle strata* (Frangakis & Rubin, 2002; White, 2005). In this thesis, we follow the work of Imbens & Rubin (1997) and refer to this new variable as the *compliance type* and denote it by U.

The definition of *compliance type* given by Imbens & Rubin (1997) is as follows:

1. *Compliers (c)* are those individuals who receive the treatment when they are assigned to the treated group. They would not receive any treatment when assigned to the control group. That is, compliers always comply with their assigned treatment Z_i.

2. *Always-takers (a)* are those individuals who always receive the treatment regardless which groups they are assigned to. That is, always-takers only comply with their assigned treatment when $Z_i = 1$.

3. *Never-takers (n)* are individuals who would never receive the treatment regardless which group they are assigned to. That is, never-takers only comply with their assigned treatment when $Z_i = 0$;

4. *Defiers (d)* are individuals who receive the treatment when they are assigned to the control group and would not receive the treatment when they are assigned to the treatment group. That is, defiers never comply with their assigned treatment $Z_i = z$ no matter whatever value Z_i has.

Using the symbol of potential received treatment $\mathbf{D}_i = [D_i(1), D_i(0)]$, these four compliance types can be written mathematically as follows,

$$U_i = \begin{cases} c, & \text{(i.e. subject } i \text{ is a complier)} & \text{if } D_i(1) = 1 \text{ and } D_i(0) = 0; \\ n, & \text{(i.e. subject } i \text{ is a never-taker)} & \text{if } D_i(1) = 0 \text{ and } D_i(0) = 0; \\ a, & \text{(i.e. subject } i \text{ is an always-taker)} & \text{if } D_i(1) = 1 \text{ and } D_i(0) = 1; \\ d, & \text{(i.e. subject } i \text{ is a defier)} & \text{if } D_i(1) = 0 \text{ and } D_i(0) = 1. \end{cases} \tag{1.7.2}$$

This compliance type variable, U_i, allows us to classify a population into four subpopulations by the four categories of individuals' compliance behaviour, that is *compliers* ($U_i = c$), *always-takers* ($U_i = a$), *never-takers* ($U_i = n$) and *defiers* ($U_i = d$).

In the literature, the fourth compliance type, *defier* ($U_i = d$), is often omitted, since there is no reason to suppose that individuals who agree to participate in a trial will necessarily always counter their experimenters. To avoid this extreme situation, a very common assumption named Monotonicity (Angrist et al., 1996; Imbens & Rubin, 1997; Frangakis & Rubin, 1999) is given, which supposes that the component of the received treatment $\mathbf{D}_i$ is a monotone function with respect to Z_i, that is $D_i(1)$ and $D_i(0)$ have a relationship as follows,

$$D_i(1) \geq D_i(0). \tag{1.7.3}$$

Instead of using the terminology of Monotonicity, some authors directly assume, however, that there are strictly no defiers in the study (Baker, 1998; Baker & Kramer, 2005).

1.8 Aim of this research

The purpose of this research is to develop a new approach for causal inference in survival analysis, that is based on the Frangakis-Rubin (F-R) method (Frangakis & Rubin, 1999). Our approach considers noncompliance and missing outcomes (censoring) together, which is unlike previous research. Our focus is the analysis of time-to-event data, rather than binary or dichotomous outcomes or normal distributed data, as in the literature of this area to date (Mealli et al., 2004; Peng et al., 2004; O'Malley & Normand, 2005). This is the first study to develop a likelihood-based (EM) approach for survival data for RCT's which accommodates both missing outcomes and noncompliance simultaneously.

The research of this thesis constructs survival modelling within the potential outcomes approach (Greenland, 2004) for causal inference, ultimately constructing a class of causal survival models which appropriately model the time dynamic nature of causal processes, whether in medicine or in other RCT survival analytic scenarios.

In this thesis, a class of extended parametric survival models will be established by following assumptions given in Frangakis & Rubin (1999). These can be applied to the analysis of the efficacy of a treatment in RCTs with noncompliance and allows also for an ignorable or non-ignorable censoring mechanism.

1.8.1 Main problems

To achieve the goals stated above, we address the following issues; namely that of

1. not fully observable potential-outcome vectors;

2. unobservable compliance type; and

3. an ignorable or non-ignorable censoring mechanism.

These three issues are discussed in brief below.

Not fully observable potential-outcome vectors

For the potential outcomes vector $\mathbf{Y}_i = [Y_i(0), Y_i(1)]$, either $Y_i(0)$ or $Y_i(1)$ is observed. In practice, each individual would only be assigned to one of the treatment groups and receive one treatment only in a RCT. This means that in a RCT, each individual's potential outcomes vector is not fully observable; instead, only one component would be observed if the individual is not censored. Using the notation given in Section 1.7.1, the potential outcomes vector $\mathbf{Y}_i$ can then be written as $\mathbf{Y}_i = [Y_i^{obs}, Y_i^{mis}]$.

Unobservable compliance type

By the definition of *compliance type* (Section 1.7.2), we require knowledge of the full vector $\mathbf{D}_i = [D_i(1), D_i(0)]$ to determine individual i's compliance type $U_i = u$. However, $\mathbf{D}_i$ is not fully observable as stated in Section 1.7.1. Therefore, compliance type (U_i) is an unobserved variable. This is a common challenge in noncompliance analysis (Imbens & Rubin, 1997; Frangakis & Rubin, 1999; Yau & Little, 2001; Peng et al., 2004; O'Malley & Normand, 2005).

Non-ignorable censoring mechanism

Apart from the two problems stated above (Section 1.8.1), which accompany the fundamental problem in causal inference (Section 1.4.4), that only one of the potential outcomes is observed in a RCT (Holland, 1986)), another challenge arises in time-to-event study due to censoring. Because of the presence of noncompliance, we have to consider the impact of both noncompliance and censoring together. This implies that the traditional assumption in survival analysis, which requires an independent censoring or ignorable censoring mechanism, may not hold (Baker, 1994). Instead, the assumption of *latent ignorability* (Frangakis & Rubin, 1999) has to be assumed. However, the latter needs to be specified in a form to suit time-to-event studies,

as censoring is slightly different to the usual case of non-response. In fact, censoring can be viewed as a special form of missing data. This feature leads to more complexities in the estimation of the *causal effect* in time-to-event RCTs.

In this thesis, we commence with the original potential outcomes framework and combine it with features of RCTs for time-to-event studies, such as censoring and noncompliance. Thereby we create a class of parametric potential-outcome survival models (PPO-survival models), pertinent to standard intention-to-treat (ITT) analyses. Recall that the ITT approach traditionally only provides an estimator of the efficacy of assigned treatment, rather than of the efficacy of the actual received treatment. As mentioned in Section 1.6, when noncompliance occurs, it is well known the ITT (Intention-to-treat) is biased estimator of effect of received treatment. Furthermore, when non-ignorable non-response occurs, e.g. non-ignorable censoring in time-to-event studies, the ITT may also be biased for effect of assigned treatment (Frangakis & Rubin, 1999).

Because of the presence of censoring, which is the major feature in time-to-event data, the new definitions of the Average of Causal Effect (ACE) and the Complier Average of Causal Effect (CACE) are given in the context of this thesis (Section 3.3). This, as derived, is based on the survival function, a traditional and well-used quantity in survival analysis; ours is a novel approach to causal effect estimation.

1.9 Overview of the thesis

This thesis has a structure as follows: Basic concepts from several topics, including causal inference, time-to-event studies, randomised RCTs and noncompliance are introduced in Chapter 1. Chapter 2 concentrates on a literature review for potential-outcome models. Chapter 3 describes the distinctions of time-to-event analysis from general data analysis, where the two important quantities for causal inference, in terms of Average Causal Effect (ACE) and Complier Average Causal Effect (CACE), have been redefined to accommodate for survival outcomes. Chapter 4 addresses different types of censoring mechanisms with their related likelihood function structures.

Theoretical parametric potential-outcome survival models, (PPO-survival models for short), are given in Chapter 5. The PPO survival models (I) are established for causal inference on time-to-event data, which are associated with the case where noncompliance is non-ignorable and the censoring mechanism is ignorable or non-ignorable. All the theoretical models established in Chapter 5 are illustrated in Chapter 6 via their specified forms on the HIP data (Baker, 1998; Shapiro et al., 1988), where Weibull or log-normal distributions are assumed for both the failure and censoring time distributions. The PPO-survival models (II), which extend the PPO-survival models (I) to incorporate pre-treatment covariates, are derived theoretically and illustrated briefly in Chapter 7. The discussion and conclusions are given in Chapter 8.

Overall, the first two chapters are background material related to the main topic addressed in this thesis. Chapters 3 and 4 provide new ideas: the foundational concepts for establishing our PPO-survival models (I) and (II). Chapter 5 and Chapter 7 are the main parts of this thesis, original contributions to the potential outcomes, survival analysis and potential outcome literature; where we address the creation of our PPO-survival by way of models (I) and (II). Chapter 6 pertains to the application of our new PPO-survival models (I). This is the first time PPO-survival models (I) have been applied. In addition, Section 6.4 provides the application of the Frangakis-Rubin model (Frangakis & Rubin, 1999) on the HIP data, which is also an original work as we are not aware of any relevant previous work in the literature.

Chapter 2

Potential-outcome models

Having given basic concepts which relate to the topics of causal inference or time-to-event studies in Chapter 1, this chapter concentrates on defining the causal effect in the potential-outcomes framework (Section 2.1) and also reviews potential-outcome models in the current literature (Section 2.3). In addition, further discussion about potential-outcome models is given in Sections 2.4 and 2.5, which include the limitations of potential-outcome models and the current state of development of them in the literature. The last section of this chapter gives the summary.

2.1 Definition of causal effects

In Section 1.4.3, the causal effect, a quantity for measuring causal influence, was roughly described as a contrast between potential outcomes (Rubin, 1974, 1978). In fact, this contrast can be obtained in two ways: one is a difference and the other is a ratio. In other words, the causal effect measure has more than one form. The one based on the ratio is termed the *relative causal effect*. This is distinct from *causal effect*, which is obtained via the difference of two potential outcomes. Furthermore, the causal effect can also be defined at both an individual and a population level. In this section, we give the definitions of *causal effect* and *relative causal effect* for both individual (Section 2.1.1) and population (Section 2.1.2) levels. Recall that in the potential outcomes framework, the outcome of interest in a two-arm RCT can be written as a vector, that is $\mathbf{Y}_i = [Y_i(0), Y_i(1)]$. Here $Y_i(0)$ and $Y_i(1)$ denote the outcome of interest corresponding to the two different levels of assigned treatment, i.e. $Y_i(0)$ is for the control (untreated), and $Y_i(1)$ is for the assigned treatment.

2.1.1 Individual level causal effects

At the individual level, the *causal effect* is called the *individual causal effect*, which is the difference between two potential outcomes of an individual and is denoted by ICE_i.

Given the potential outcomes vector $\mathbf{Y}_i = [Y_i(0), Y_i(1)]$ for individual i, ICE_i can then be expressed as follows:

$$ICE_i = Y_i(1) - Y_i(0). \tag{2.1.1}$$

Similarly, the *relative causal effect* at an individual level, namely the *individual relative causal effect*, denoted by IRE_i, is the ratio between the two potential outcomes (if they all are positive). IRE_i can be expressed as follows:

$$IRE_i = \frac{Y_i(1)}{Y_i(0)} \qquad \text{for } Y_i(0) > 0 \text{ and } Y_i(1) > 0. \tag{2.1.2}$$

Note that in equation (2.1.2) above IRE_i is undefined when $Y_i(0) = 0$.

2.1.2 Population level causal effects

At the population level, there are two forms of so-called *Averaged Causal Effects* (ACE): one from the individual causal effects, which is based on ICE_i (denoted by ACE), and the other from the individual relative causal effects, based on IRE_i (denoted by RCE). These can be written as

$$ACE = E(Y_i(1)) - E(Y_i(0)), \tag{2.1.3}$$

and

$$RCE = \frac{E(Y_i(1))}{E(Y_i(0))}, \tag{2.1.4}$$

where $E()$ denotes an expectation over all the population of treated or control individuals.

By following the property of expectation, the following equality always holds:

$$E(Y_i(1)) - E(Y_i(0)) = E(Y_i(1) - (Y_i(0)). \tag{2.1.5}$$

Hence the ACE can be re-expressed as follows:

$$ACE = E(ICE_i). \tag{2.1.6}$$

The relationship given in equation (2.1.6) implies that a potential-outcome model for estimating the ACE (Average Causal Effect) is often related to the ICE (Individual Causal Effect), we name this type of model as ICE based potential-outcome model, which can be seen in the literature on causal inference. In fact, ACE has attracted much attention in the last decades for general data (Angrist et al., 1996; Imbens & Rubin, 1997; Yau & Little, 2001; Peng et al., 2004; O'Malley & Normand, 2005), whereas the RCE is less tractable for time-to-event data because $\frac{E(Y_i(1))}{E(Y_i(0))} \neq E(\frac{Y_i(1)}{Y_i(0)})$, which implies $RCE \neq E(IRE_i)$.

2.2 Role of randomisation

In Section 1.5, we stated that the treatment and control groups in a RCT are able to be comparable because they can be viewed as two samples, which are drawn randomly and independently from the same population. This implies that the outcome variable $\mathbf{Y}_i = [Y_i(1), Y_i(0)]$ is independent of the assigned treatment Z_i. That is $[Y_i(1),\ Y_i(0)] \perp Z_i$. So the conditional distributions of the potential outcomes $Y_i(1)$ or $Y_i(0)$, given the assigned treatment $Z_i = z$ (for $z = 1$ or 0), can be regarded as the same. That is:

$$\begin{aligned} Pr(Y_i(1)|Z_i = 1) &= Pr(Y_i(1)|Z_i = 0) = Pr(Y_i(1)), \\ Pr(Y_i(0)|Z_i = 1) &= Pr(Y_i(0)|Z_i = 0) = Pr(Y_i(0)), \end{aligned} \tag{2.2.1}$$

where $Pr()$ denotes the probability distribution. This leads to the following equalities hold:

$$\begin{aligned} E(Y_i(1)|Z_i = 1) &= E(Y_i(1)|Z_i = 0) = E(Y_i(1)), \\ E(Y_i(0)|Z_i = 1) &= E(Y_i(0)|Z_i = 0) = E(Y_i(0)), \end{aligned} \tag{2.2.2}$$

where $E()$ denotes the expectation that takes over all the population of treated or control individuals.

Therefore equation (2.1.3) can be replaced by the following:

$$ACE = E(Y_i(1)|Z_i = 1) - E(Y_i(0)|Z_i = 0). \tag{2.2.3}$$

If we write $\mu_1 = E(Y_i(1))$ and $\mu_0 = E(Y_i(0))$, then via equation (2.2.2), randomisation implies $\mu_1 = E(Y_i(1)|Z_i = 1)$ and $\mu_0 = E(Y_i(0)|Z_i = 0)$, which can be estimated in randomised trials, via

$$\begin{aligned} \hat{\mu}_1 &= \frac{\sum_{i\in\{i:Z_i=1\}}^{N_1} Y_i(1)}{N_1}; \\ &\text{and} \\ \hat{\mu}_0 &= \frac{\sum_{i\in\{i:Z_i=0\}}^{N_0} Y_i(0)}{N_0}, \end{aligned} \tag{2.2.4}$$

where N_1 and N_0 denote the number of subjects in the treatment group and control group respectively. The ACE, *average causal effect*, can then be estimated straightforwardly as follows

$$\widehat{ACE} = \hat{\mu}_1 - \hat{\mu}_0. \tag{2.2.5}$$

Similarly, the RCE can be estimated by

$$\widehat{RCE} = \frac{\hat{\mu}_1}{\hat{\mu}_0} \quad \text{if } \hat{\mu}_0 \neq 0. \tag{2.2.6}$$

Equations (2.2.5) and (2.2.6) thus provide a possible way to solve the fundamental problem in casual inference (Section 1.4.4), where randomisation plays an important role.

2.3 Potential-outcome models

Since Neyman in 1923 first proposed the concept of the potential outcomes framework, which in itself is the first example of a *potential-outcome model* (Greenland, 2004), potential-outcome models have been attractive, especially over the last decade (Angrist et al., 1996; Imbens & Rubin, 1997; Frangakis & Rubin, 1999; Hirano & Imbens, 2000; Graham, 2000; Barnard et al., 2003; Rubin, 2004; O'Malley & Normand, 2005).

Neyman's model is a classic and formal statistical model for causal inference (Greenland, 2004) and is also called the Rubin Causal Model (RCM) in the literature (Holland, 1986; Rubin, 2004; Angrist et al., 1996; Frangakis & Rubin, 1999; Hirano & Imbens, 2000; Graham, 2000; Barnard et al., 2003; Rubin, 2004; O'Malley & Normand, 2005), even though the RCM refers to the Neyman model which was proposed earlier. However, apart from the RCM model, in the literature there are still some other potential-outcome models such as the structural nested failure time model (SNFTM) (Robins (1998), Robins (1998a)); the marginal structural model (MSM) (Robins (1999), Robins (1999a), Robins et al. (1999b), Hernan et al. (2000, 2001)), the component equations in structural-equation model (SEM) (Pearl, 1995, 2000; Jeffrey, 2007) and the causal diagrams model (Pearl, 1995).

In this thesis, we class potential-outcome models into two types by their applied data type; one is for general data causal inference models that are based on *Average Causal Effect* (ACE) or *Complier Average Causal Effect* (CACE) analysis, and the other is for time-to-event data causal inference models which are mostly derived from a semi-parametric survival model but have not been related to ACE estimating yet. We review the former class of potential-outcome models in Section 2.3.1, where they are referred as ACE or ICE based potential-outcome model. Another class of potential-outcome models, which we called the semi-parametric potential-outcome models, are reviewed in Section 2.3.2.

2.3.1 For general data analysis

According to the current literature, the potential-outcome models for estimating ACE or CACE, which are the ICE based potential-outcome models, have been applied only on the general data (Angrist et al., 1996; Imbens & Rubin, 1997; Hirano & Imbens, 2000; Graham, 2000; Barnard et al., 2003; Rubin, 2004; O'Malley & Normand, 2005). Most of these models proposed in the papers mentioned above are under the assumption of the *stable unit treatment value* (SUTVA), which was first given by Rubin (1986) and assumed that individuals' potential outcomes are unrelated to the treatment status of other individuals (Graham, 2001; Peng et al., 2004; Mattei & Mealli,

2007); where the definition of causal effect at the individual level (ICE) was given exactly by equation (2.1.1) (Holland, 1986) but not equation (2.1.2), as general data may have zero or negative values.

Under randomisation, especially in *intention-to-treat* analysis (Section 1.6), the causal effect at the population level (ACE) can be specified by equation (2.2.3), and the ACE, can easily be estimated by equation (2.2.5) (Holland, 1986).

After adoption of an instrumental variable technique and via two associated assumptions, the *exclusion restriction* (equation (1.7.1)) and *monotonicity* (equation (1.7.3)), Angrist et al. (1996) extended these ACE based models to deal with the critical problem of evaluating the treatment effect in a randomised trial when noncompliance occurred, but without consideration of non-responses.

Imbens & Rubin (1997) then used the compliance type covariate, whereby the ACE (equation (2.2.3)) can further be written as the difference between the potential outcomes of compliers, namely *complier average causal effect* (CACE) in a noncompliance scenario, that is

$$CACE = E(Y_i(1)|Z_i = 1, U_i = c) - E(Y_i(0)|Z_i = 0, U_i = c). \tag{2.3.1}$$

Furthermore, given the *exclusion restriction* assumption (Imbens & Rubin, 1997)(equation (1.7.1)), it leads to the following equations:

$$\begin{aligned} E(Y_i(1)|Z_i = 1, U_i = a) &= E(Y_i(0)|Z_i = 0, U_i = a); \\ E(Y_i(1)|Z_i = 1, U_i = n) &= E(Y_i(0)|Z_i = 0, U_i = n). \end{aligned} \tag{2.3.2}$$

The *average causal effect* (ACE) can be expressed as:

$$ACE = \pi_c CACE \tag{2.3.3}$$

where π_c is the probability of compliers ($U_i = c$), i.e. $\pi_c = Pr(U_i = c)$ (Imbens & Rubin, 1997).

Recall that because of randomisation, the covariate of compliance type is independent of the assigned treatment Z_i, that is $U_i \perp Z_i$. Hence the following

equation holds,

$$Pr(U_i = c|Z_i = 1) = Pr(U_i = c|Z_i = 0) = Pr(U_i = c). \tag{2.3.4}$$

Hence $\pi_c = Pr(U_i = c|Z_i = 1)$ can be estimated by $\hat{\pi}_c = \frac{N_{11}}{N_1}$, where N_1 denotes the number of individuals in the treated group and N_{11} is the number of individuals who do actually receive the treatment in the treated group (Imbens & Rubin, 1997; Baker, 1998; O'Malley & Normand, 2005).

By adding in the *latent ignorability* assumption, which assumes that the potential outcomes and associated potential non-response indicator are independent (for each level of the latent compliance types), Frangakis & Rubin (1999) developed the extended potential-outcome model to estimate the causal effect of treatment, which accommodated both noncompliance and non-response mechanisms. As such, they introduced a more plausible extended potential-outcome model which appeared in the literature as the so-called Frangakis-Rubin model (F-R model) (Mealli et al., 2004).

Amongst the extensive body of literature on noncompliance, most papers address noncompliance as missing covariates. This is because the compliance type $U_i = u$ is considered to be an unobservable variable, at least for those individuals in the control group. Hence the compliance type is often regarded as a missing covariate (O'Malley & Normand, 2005; Frangakis & Rubin, 1999; Baker, 1998). Some recently published papers, however, do consider noncompliance and subsequent non-response together (Frangakis & Rubin, 1999; Baker, 1998; Hirano & Imbens, 2000; Barnard et al., 2003; Peng et al., 2004; Mealli et al., 2004; O'Malley & Normand, 2005). This is partially relevant for time-to-event studies, which often involve censoring and hence missing outcomes (of kinds).

The Frangakis-Rubin (F-R) model (Frangakis & Rubin, 1999; Mealli et al., 2004) is the first potential outcomes model to consider both noncompliance and non-response mechanisms together. For intention-to-treat (ITT) analysis, a standard approach can be biased, in some settings, because of the need to omit subjects with missing outcomes and ignore an individual's compliance information. Frangakis & Rubin (1999) constructed a new estimation procedure which was designed to address all-or-none compliance

and allows for missing outcomes. In this way, the ITT effect could maintain good randomisation-based properties under more plausible conditions, namely that of non-ignorable noncompliance and of non-ignorable missing outcomes. Frangakis & Rubin (1999) also allow for the *compound exclusion restriction* to hold on the effect of assignment and for *latent ignorability* to be valid for the missing data mechanism. The Frangakis-Rubin (F-R) model supposed that individuals who were assigned to the control group did not have any chance to receive the intervention treatment. This assumption implies that no *always-takers* exist in the randomised trial. In this case, compliance types are thus reduced and can be indicated by a binary variable. The two key concepts of *compound exclusion* and *latent ignorability* are also relevant in even more complicated settings, such as right censoring of a time-to-event outcome (Frangakis & Rubin, 1999; Klein, 1997).

For the analysis of data from an extended framework experiment with the complexities of a large amounts of noncompliance, a binary treatment regime, binary encouragement and background covariates, Hirano & Imbens (2000) adopted a Bayesian approach for inference and sensitivity analysis. They adopted an instrumental variable (IV) approach to link intention-to-treat effects with the treatment effects.

Follmann (2000) proposed two methods: a propensity score (PS) approach and a counterfactual (CF) approach (which is specific to the Weibull distribution). Follmann (2000) examined how the effect of treatment varies with compliance to treatment via a pseudo-likelihood method to derive the maximum likelihood estimators (MLEs).

Based on the maximum likelihood (ML) approach and Bayesian inferential methods for CACE, a model for inference was developed by Yau & Little (2001), which allowed for the inclusion of baseline covariates and handled missing data in repeated measures. Yau and Little's model (Yau & Little, 2001) was applied to a randomised trial of job training intervention for unemployed workers. The results show that job training intervention significantly reduced depression for high-risk compliers (up to six months post-intervention) but not for low-risk compliers (Yau & Little, 2001).

In a two-armed randomised trial to compare the efficacy of two antipsychotics for adults with refractory schizophrenia, which also assumed all-or-

none compliance and allowed for subsequent non-response, O'Malley & Normand (2005) proposed a maximum likelihood estimator (MLE) of the causal effect of assigned treatment, which was based on a latent compliance type covariate. The latter described the behavior of a subject under all possible assigned treatments. This characterized the missing data mechanism as in the Frangakis-Rubin (F-R) model (Frangakis & Rubin, 1999) and used the EM algorithm for parameter estimation. The authors claimed that simulated results showed that the MLE (for normal outcomes) compares favourably to other methods such as the standard intention-to-treat (ITT) estimator, the method-of-moments (MOM) estimator (under both normal and non-normal data), and accommodated for departures from the *latent ignorability* and *compound exclusion restriction* assumption (O'Malley & Normand, 2005).

Research on the efficacy of the school choice scholarship foundation programme in New York (Barnard et al., 2003) revealed threats to valid estimates of experimental effects that exist in the presence of noncompliance and missing data. This occurred even when the goal was to estimate simple intention-to-treat (ITT) effects. A model with covariates was developed by Barnard et al. (2003) which accommodated for these complications and was based on the general framework of *principal stratification*. Barnard et al. (2003) created a better solution to allow for both noncompliance and missing data. *Principal stratification* is an approach proposed by Frangakis & Rubin (2002). In brief, the *principal stratification*, with respect to a post-treatment variable, is a cross-classification of individuals defined by the joint potential values of the given post-treatment variable under each of the treatments being compared (Frangakis & Rubin, 2002).

In a study of job search intervention for unemployed workers, Peng et al. (2004) proposed an extended general location model to estimate the CACE from data with both noncompliance and missing data in the outcome of interest and also in baseline covariates. Models for both continuous and categorical outcomes, and ignorable and latent ignorable missing data mechanisms were developed by Peng et al. (2004). Inference for these models was based on the EM algorithm and on Bayesian MCMC methods. Simulation results were used to investigate sensitivity to model assumptions and to assess the influence of the missing data mechanism (Peng et al., 2004).

Later, Mealli et al. (2004) proposed a new missing data model that, like the Frangakis-Rubin model, is specially suited for models with instrumental variables, but under different substantive assumptions. The new model has been applied on analysis data from a randomised trial of breast self-examination (BSE). Mealli et al. (2004) also claimed that their new model appears to be the most plausible of the three models, *the standard model for missing data* (missing at random), the Frangakis-Rubin (F-R) model and the author's model, where the model of missing at random appears to be the least plausible.

We now summarize the main assumptions involved in the above approaches:

1. The stable unit treatment value assumption (SUTVA) (Imbens & Rubin, 1997; Peng et al., 2004; Mattei & Mealli, 2007): the potential outcomes of each individual are unrelated to the treatment status of other individuals.

2. Random assignment: each individual has been assigned randomly to treatment group.

3. Exclusion restriction: the assigned treatment has an effect on the outcome only through its effect on the received treatment.

4. Monotonicity: the actual received treatment $D_i(Z_i)$ is a monotone function with respect to the potential treatment levels (1 for being treated and 0 for being in the control), that is $D_i(1) \geq D_i(0)$.

5. Latent ignorability: the potential outcomes and associated potential non-response indicator are independent for each level of the latent compliance types.

In general, most of methods developed from the ACE or CACE based potential-outcome models listed above, such as those proposed by Angrist et al. (1996); Imbens & Rubin (1997); Yau & Little (2001); Barnard et al. (2003); Mealli et al. (2004); Peng et al. (2004) and O'Malley & Normand (2005), do not deal directly with time-to-event data. Some studies pertain

to binary outcomes (Imbens & Rubin, 1997; Yau & Little, 2001; Barnard et al., 2003; Mealli et al., 2004) and others to general data that are assumed to follow a normal distribution (Peng et al., 2004; O'Malley & Normand, 2005).

All these approaches may be viewed as potential-outcome models for general data in noncompliance studies, where the focus is the *complier average causal effect* (CACE). The CACE has attracted much attention in the last decade (Rubin, 1998; Yau & Little, 2001; Jo, 2002a,b; Dunn et al., 2003; Mealli et al., 2004; Peng et al., 2004). However, to date, CACE analysis has not been developed for time-to-event studies. This is one of the aims of this thesis.

2.3.2 For time-to-event data analysis

In time-to-event studies, as noted in Chapter 1, the outcome of interest is defined as the event time, which usually takes positive values and does not follow a normal distribution. As the event time is the outcome of interest, it is also highly related to the occurrence of censoring. Hence most of the potential-outcome models described in Section 2.3.1 are not suitable to time-event-study scenarios as they are mainly applicable only to general data. This thesis will develop potential-outcome models which are based on the RCM, and also consider noncompliance and subsequent non-response together. These new potential-outcome models for causal inference are applicable to RCTs with noncompliance and time-to-event studies. (Details are given Chapter 5 and Chapter 7).

Before addressing our potential-outcome models, we will review some potential-outcome models that were recently proposed for time-to-event studies.

Amongst the extensive body of literature on potential-outcome models for time-to-event data analysis, the following models could plausibly be used: the structural nested failure time model (SNFTM) (Robins & Tsiatis, 1991; Robins, 1998,a; Vandebosch et al., 2005; Hernan et al., 2005), the marginal structural Cox Proportional Hazards model (Hernan et al., 2000a; Loeys & Goetghebeur, 2003; Gong et al., 2005) and the method of inverse probability

of censoring weighted log-rank tests (Robins & Finkelstein, 2000).

The common approach underlying these models is the accelerated failure time (AFT) model (p.46, Klein (1997), or p.64, Cox & Oakes (1984)), where the potential outcomes, if contrary to the fact (unobservable), can be expressed as a product of the observable time $T_i(Z_i)$ and the accelerated factor $\exp(-\beta Z_i)$, that is

$$T_i(0) = \exp(-\beta)T_i(1), \tag{2.3.5}$$

when the individual i is assigned to the treated group ($Z_i = 1$) in a two-arm RCT. Note that here, $T_i(0)$, the survival time corresponding to the control group is contrary to the fact and can not be observed in practice.

Following the definition of causal effect for positive outcomes, namely the *individual relative causal effect* (given in equation (2.1.2)), the IRE for the survival time, the outcome of interest in time-to-event studies, has the following form:

$$IRE_i = \frac{T_i(1)}{T_i(0)} = \frac{T_i(1)}{\exp(-\beta)T_i(1)} = \exp(\beta). \tag{2.3.6}$$

Therefore, the accelerated parameter β can be interpreted as the parameter of causal effect as equation (2.3.6) has shown that $\exp(\beta)$ is the causal effect for time-to-event outcomes.

Robins & Tsiatis (1991) is the earliest paper on noncompliance analysis for survival outcomes with a view to estimating the treatment effect. In Robins & Tsiatis (1991), a class of models called rank-preserving structural failure time models (RPSFTM) were introduced with a latent failure time variable, called the treatment-free time, denoted by T_i^0 (even if the treatment-free time may not be observable). Then T_i^0 is the failure time of the *ith* individual who never received the treatment. By following the development of usual time-dependent accelerated models (p.66-67, Cox & Oakes (1984)), the latent failure time variable can then be related to the observable random variables as follows:

$$T_i^0 = \int_0^{T_i} \exp(-\beta_0 D_i(t))dt. \tag{2.3.7}$$

where $D_i(t)$ refers to a time-dependent treatment indicator, ($D_i(t) = 1$ for the *ith* individual receiving the treatment at time t and $D_i(t) = 0$ otherwise);

β_0 is the unknown parameter with a causal interpretation. The treatment-free time, T_i^0 , can be viewed as the lifetime of the *ith* individual if he/she does not received the treatment of interest in the trial. Under the assumption of that the baseline lifetime T_i^0 is independent of the assigned treatment indicator Z_i, $(T \perp Z_i)$, the survival function of the baseline life time is the same in the treated and the control group. That is:

$$Pr(T_i^0 \geq t | Z_i) = S_0(t) \quad \text{for } Z_i \in (0, 1). \tag{2.3.8}$$

This implies that all individuals will have a $T_i^0 = T^0$, whether or not the actually receive the treatment.

Standard nonparametric methods can be applied to test the null hypothesis $\beta_0 = 0$. Robins & Tsiatis (1991) assumed that no censoring prior to the end of follow-up exists, which means that this model is covered by the assumption of noninformative censoring. We can make this assumption because no censoring prior to the end of follow-up implies that censoring time is a constant rather than a random variable; no censoring mechanism needs to be considered. Robins & Tsiatis (1991) also supposed that they had a rank-preserving model, which means that individuals keep their order of failure time when they receive the same treatment. For instance, suppose that any two individuals, i and j, both received a particular treatment. If individual i fails before individual j, then for any other treatment they both receive, individual i would fail before individual j (Robins & Tsiatis, 1991).

Mark & Robins (1993) proposed an approach to analyzing randomised trials without ignoring information on postrandomization compliance, which uses Robins and Tsiatis' method and takes advantage of postrandomization smoking history without requiring untenable assumptions about the comparability of compliers and noncompliers in the data from the Multiple Risk Factor Intervention Trial (MRFIT).

More generally, the model is also known as the so-called the structural nested failure time model (SNFTM) (Robins, 1998,a; Vandebosch et al., 2005; Hernan et al., 2005), a generalization of the strong accelerated failure time (AFT) model (p.46, Klein (1997), or p.64, Cox & Oakes (1984)), in which it is supposed that the assigned treatment (Z_i) may vary and the actual

survival time interval, $\mathcal{L}_{\mathcal{S}} = [0,\ T_i(z)]$, (0 denotes the start point and $T_i(z)$ represents the point of time when individual i in the $Z_i = z$ treatment group experienced the event of interest) is divided into K successive intervals of length $\triangle t_1, \ldots, \triangle t_K$. The potential survival time, $T_i(0)$, corresponding to the control group, can then be written as follows:

$$T_i(0) = \sum_k \exp(-\beta z_k) \triangle t_k; \tag{2.3.9}$$

when the treatment history is continuous, $z(t)$ in, equation (2.3.9) can then be replaced by the following equation:

$$T_i(0) = \int_{\mathcal{L}_{\mathcal{S}}} \exp(-\beta z(t)) dt. \tag{2.3.10}$$

In the marginal structural Cox proportional hazards (PH) model (Hernan et al., 2000; Loeys & Goetghebeur, 2003; Gong et al., 2005), the Cox proportional hazards model (p. 230, Klein (1997), or p.70, Cox & Oakes (1984)), which is a standard semi-parametric method, was adopted rather than the accelerated failure time model (p.46, Klein (1997) or p.64, Cox & Oakes (1984)). The subject of interest in the marginal structural Cox proportional hazard (PH) model is the hazards rate function (Section 3.1.2) rather than survival time itself. Here the hazard rate functions, corresponding to the observed treatment level $Z_i = 1$ and $D_i(1) = 1$ and the unobservable potential treatment level, $Z_i = 0$ and $D_i(0) = 1$ (counterfactual), are denoted by $\lambda_{11}(t)$ and $\lambda_{10}(t)$ respectively.

By following Cox's proportional hazards model (p.230, Klein (1997) or p.70, Cox & Oakes (1984)), the two hazard rate functions have a relationship as follows:

$$\lambda_{11}(t) = \lambda_{10}(t) \exp(\psi), \tag{2.3.11}$$

where $\exp(\psi)$ denotes the causal proportional hazards effect within the treatable subpopulation (Loeys & Goetghebeur, 2003).

Korhonen & Laird (1999) discussed different methods, which include the *intention-to-treat* approach, the *as-treated* approach and the g-estimation approach, for estimating the effect of treatment actually received in a lon-

gitudinal placebo-controlled trial with noncompliance. All these models are based on the Cox proportional hazard model (p.230, Klein (1997), or p.70, Cox & Oakes (1984)), where noncompliance is treated as non-ignorable and the censoring mechanism has been ignored.

Korhonen (2000) addressed several accelerated failure time models for dealing with the noncompliance issue in randomised clinical trials. Even though the author did not claim that the mechanism of censoring was ignorable, at least they are not under the *latent ignorability* assumption (Frangakis & Rubin, 1999).

Robins & Finkelstein (2000) proposed a new method: the so-called Inverse Probability of Censoring Weighted (IPCW) log-rank test for estimating the effect of treatment with noncompliance and a non-ignorable censoring mechanism. Here a fundamental assumption is assumed, namely no unmeasured confounders of censoring (Robins, 1997) or, equivalently, the assumption of sequential ignorability of censoring (Robins & Finkelstein, 2000). It is expressed in equation (2.3.12), which means that conditional on the assigned treatment Z_i and recorded history $H_i(t)$ of all associated time-dependent covariates $X_i(t)$, where $H_i(t) = \{X_i(u); 0 \leq u \leq t\}$, the cause-specific hazard rate of censoring $\lambda_C(t)$ (at time t, $C_i = t$) does not further depend on the possibly unobserved failure time T_i, i.e.

$$\lambda_C(t|H(t), Z, T, T > t) = \lambda_C(t|H(t), Z, T > t). \tag{2.3.12}$$

The usual assumption of independent censoring is that, given the treatment group $Z_i = z$, the cause-specific hazard of failure time $T_i = t$ (for individuals at risk) denoted by $\lambda_T(t|Z, C \geq t)$ is the marginal hazard of failure time $T_i = t$, $\lambda_T(t|Z)$ (Robins & Finkelstein, 2000). This can be expressed as:

$$\lambda_T(t|Z, C \geq t) = \lambda_T(t|Z). \tag{2.3.13}$$

With respect to the study of causal inference for survival data obtained from cluster randomised trials, Loeys et al. (2001) proposed that structural models can be employed to account for post-randomisation exposures, but that structural models should not ignore clustering. This was illustrated using a Vitamin A trial for the prevention of infant mortality in the rural

plains of Nepal (Loeys et al., 2001). Marginal modelling and random effects models were used to adapt structural estimators to account for clustering (Loeys et al., 2001).

In survival analysis of standard proportional hazards (PH) models that traditionally follow the intention-to-treat (ITT) principle, Loeys & Goetghebeur (2003) derived an estimating equation for the Compliers PROPortional Hazards Effect of Treatment (C-PROPHET) in an analysis of data from a clinical trial of cancer patients with liver metastases. The jackknife method was used for bias correction and variance estimation, while potential treatment compliance was considered as unobserved in the control arm. Loeys & Goetghebeur (2003) clearly claimed that their proposed C-PROPHET model was under the noninformative censoring assumption.

More recently, Hernan et al. (2005) used a nested structural accelerated failure time model to estimate the total causal effect of a highly active antiretroviral therapy (HAART) on the time to AIDS (or death) among human immunodeficiency virus (HIV) infected participants. This was a part of the Multicenter AIDS Cohort and Women's Interagency HIV Studies (Hernan et al., 2005).

In the analysis of causal effects of observed exposures in randomised trials, structural accelerated failure time models may thus be seen to be selective i.e. reliant on unmeasured prognostic factors, when the actual exposure levels are not completely controlled for the effect of such exposures (on a right censored survival outcome). Note also that Vandebosch et al. (2005) derived consistent randomisation-based estimators to allow estimation of the structural effect of an experiment.

Note that almost all the potential-outcome models listed above are based on estimating the hazard rate. It is noteworthy that the structural accelerated failure time models (Robins & Tsiatis, 1991; Vandebosch et al., 2005; Hernan et al., 2005) provide consistent estimates of causal effects when unmeasured confounding and model misspecification are absent, thereby giving bias in the standard time-dependent accelerated failure time (AFT) model (p.66, Cox & Oakes (1984)).

The approaches mentioned above, for noncompliance analysis and for survival outcomes in imperfect randomised trials, are mostly based on semi-

parametric (Korhonen & Laird, 1999; Korhonen, 2000; Robins & Finkelstein, 2000; Loeys & Goetghebeur, 2003) or non-parametric (Robins & Tsiatis, 1991) methods in survival analysis. They all depend on certain specific assumptions for the underlying censoring mechanism.

2.4 Limitation of current potential-outcome models

In this chapter, potential-outcome models have been categorized by the type of outcome of interest: either 1) general data, or 2) time-to-event data. The first type (general data) is related to estimation of the causal effect for non-time-to-event data, where the definition of casual effect follows the concept of *average causal effect* (ACE). ACE is defined as the difference between the components' mean of the potential outcomes vector (equation (2.1.3)), its related model approaches include both parametric and non-parametric models. The second type (time-to-event data) is based primarily on semi-parametric methods and focuses on survival outcomes. Here, the definition of casual effect follows the concept of *relative causal effect* (Graham, 2001). Among the approaches reviewed in Section 2.3.1, some are primarily based on the ICE at the individual level and aim at estimating *average causal effect* (ACE) or *complier average causal effect* (CACE) for the population level estimate.

In addition, most studies focus on outcomes which are binary (Sommer & Zeger, 1991; Barnard et al., 2003; Mealli et al., 2004; Robins & Rotnitzky, 2004) or normally distributed (Yau & Little, 2001; O'Malley & Normand, 2005; Peng et al., 2004), or alternatively, approach inference from a non-parametric perspective (Frangakis & Rubin, 1999). Only the approaches, proposed by Frangakis & Rubin (1999) and by Baker (1998) respectively, have concerned to survival outcomes with the noncompliance and with an assumed non-ignorable censoring mechanism. The former is based on the Kaplan-Meier estimator (Section 3.2.1), considered to be the standard non-parametric estimator in survival analysis; the latter is based on cause-specific hazard and provides a likelihood form to estimate the cost-effectiveness of a cancer screening programme (Baker, 1998). Both of these approaches (Frangakis & Rubin (1999) and Baker (1998)) assume *no defiers* (monotonicity)

and *exclusion restriction*, but only Frangakis & Rubin (1999) assume *latent ignorability*, which can viewed as a special case of informative censoring for survival outcomes. The method used in Robins & Finkelstein (2000) considers both noncompliance and informative censoring, but it is based on the semi-parametric model rather than the parametric model.

All the potential-outcome models for time-to-event data (Section 2.3.2) are based on the hazard rate and use semi-parametric methods. To the best of the author's knowledge, no survival model which considers noncompliance behaviours and censoring mechanism together, are to be found in the literature to date. Instead of this simultaneous consideration, Robins & Tsiatis (1991) and Korhonen (2000) proposed accelerated failure time models which accommodate the non-ignorable noncompliance issue; alternatively, Rotnitzky et al. (1997) proposed an approach which accommodates for non-response. In other words, among all the potential-outcome models discussed in Section 2.3, we have not found any parametric potential model that can be used to deal with problems arising from the case such as the HIP breast cancer study, which involves the following: noncompliance behaviour, non-response (censoring) and survival outcomes. This is the motivation of this research.

2.5 Direction of developing potential-outcome models

According to the literature reviewed above, most work (Robins & Tsiatis, 1991; Korhonen & Laird, 1999; Loeys & Goetghebeur, 2003; Gong et al., 2005) on potential-outcome models for time-to-event data are based on the assumption of non-informative censoring. This is equivalent to assuming independent censoring. In some trials, however, it is hard to claim that failure and censored time are independent. Note the case that has been addressed in Frangakis & Rubin (1999), where the failure and censoring time may be independent conditional on some other unobservable covariates, e.g. compliance type, for the scenario where noncompliance is an issue of concern.

In the HIP breast cancer trial, with its high proportion of non-compliers, noncompliance is not ignorable. Therefore, we cannot simply assume independent censoring. Following the assumption of *latent ignorability* (Fran-

gakis & Rubin, 1999), we consider an independent censoring mechanism conditional on compliance types (Baker, 1998; Frangakis & Rubin, 1999). In other words, censoring is likely to be informative in this situation. Frangakis & Rubin (1999) thus proposed their methodology for survival data via a nonparametric survival method which was based on the usual Kaplan-Meier estimator (Section 3.2.1) rather than one based on parametric models (Section 3.2.3).

However, with the rapid development of computer techniques, many computational methods are now available, such as the Expectation Maximization (EM) algorithm (Dempster et al., 1977), and Markov-chain Monte Carlo (MCMC) methods, which can be adopted for parametric studies with missing data. Therefore, all models can be applied to parametric survival studies with noncompliance.

The EM algorithm supplements the maximum likelihood-based method proposed by Dempster et al. (1977), and has been further developed by Louis (1982) and Meng & Rubin (1993). The EM algorithm is now widely adopted when dealing with missing data, particularly in handling unobservable variables such as compliance type variables in noncompliance studies (Peng et al., 2004; O'Malley & Normand, 2005). The approach proposed in this thesis adapts the EM algorithm as necessary for time-to-event data.

2.6 Summary

This chapter reviews the potential-outcome models for two types of data: general data and time-to-event data. These methods can be viewed as the models developed from Neyman's potential-outcome model or the RCM but which follow the different definitions of causal effect. For general data, it is based on the *individual causal effect* and for time-to-event data, it can be regarded from *relative causal effect* perspective.

After reviewing the current literature on potential-outcome models, we observe that, to date, no parametric potential-outcome model can be used to estimate causal effect for time-to-event data from RCTs with all-or-none compliance, as the case for the HIP breast cancer trial. This is precisely the motivation for doing this project.

Chapter 3

Causal effect for time-to-event data

As mentioned in Chapter 1, time-to-event data is a special type of data where *censoring* is often present and event time of the outcome of interest never follows a normal distribution. In this chapter, more details about time-to-event data are addressed. The features of time-to-event data are addressed in Section 3.1, some conventional statistical methods for survival analysis are introduced briefly in Section 3.2 and a new definition of causal effect for time-to-event data is given in Section 3.3.

3.1 Features of time-to-event data

In time-to-event studies, it would be impossible to expect that all individuals involved in the trial have available failure time, as some are likely not to experience the pre-specified event of interest during the observation period. This is a common feature of time-to-event data and is known as *censoring* (p.55, Klein (1997)). For censored individuals, their real failure time is not known; only their censoring time is observed. In the case of missed failure time, we can still gather some information about censored individuals via their observed censoring time. We need, however, knowledge about the underlying censoring mechanism involved in the given RCT.

3.1.1 Censoring types

In general, three censoring types exist: *right censoring*, *left censoring* and *interval censoring*. Right censoring is the case where an individual may experience the event of interest after the given time t; we know only that the individual is alive (not failed) up to the given time (p.56, Klein (1997). Left censoring is where an individual has experienced the event of interest prior

to the start of the study (p.62, Klein (1997)). Interval censoring is where the only information is that the event occurs within some interval of time (p.63, Klein (1997)).

Let T and C represent the failure and censoring time respectively. Then the three types of censoring can be expressed mathematically as the following:

Right censoring : $T \in (C_r,\ \infty)$ and it is known only that the failure time T is greater than the observed censoring time C_r, but exact value of failure time is unobservable.

Left censoring : $T \in (0,\ C_l)$ and it is known only that the failure time T is less than the observed censoring time C_l, but its exact value is unobservable.

Interval censoring : $T \in (C_l,\ C_r)$ and it is known only that the failure time T is less than the observed right censoring time C_r and greater than the observed left censoring C_l, but its exact value is unobservable.

For example, if individuals are right censored at time $C_i = t$, we know that their failure time would be at least greater than t, that is $T > t$ (Klein, 1997; Cox & Oakes, 1984).

Besides *censoring*, there is another feature, the so-called *truncation* issue, which may also be present in some time-to-event studies. We do not consider this concept in this thesis, as our causal effect and potential-outcome approach does not consider truncated data. In fact, the approach proposed in this thesis is based only on *right censoring*. It may will be possible to develop our methods later for other types of censoring, such as *interval censoring*. In the remainder of the thesis, when *censoring* appears, it always means *right censoring* unless otherwise specified.

For analysing time-to-event data, two quantities are possibly of greater interest than failure time itself. They are known as the *survival function* and the *hazard (rate) function* with respect to failure time t.

3.1.2 Survival function and Hazards function

The *survival function* $S(t)$ is the probability of a subject surviving after a given time point t, i.e.:

$$S(t) = Pr(T_i > t) = \int_t^{\infty} f(u)du, \tag{3.1.1}$$

where $Pr()$ denotes the probability function, T_i is the failure time of the individual i and $f()$ represents the probability density function (p.d.f.).

The *hazard (rate) function* $h(t)$ is the instantaneous rate of failure at given time t and can be written as:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T_i \leq t + \Delta t | T_i \geq t)}{\Delta t} = \frac{f(t)}{S(t)}, \tag{3.1.2}$$

where $S(t)$ is the survival function at time t, defined in equation (3.1.1) (Klein, 1997; Cox & Oakes, 1984; Kalbfleisch & Prentice, 2002); and $f(t)$ is the same as in equation (3.1.1).

3.1.3 Censoring mechanisms

There are several types of censoring schemes which lead to different likelihood functions for inference. These are delineated below.

Type I censoring

For Type I censoring, the event is observed only if it occurs prior to some pre-specified time. Censoring time may vary from individual to individual. Owing to cost or time considerations, the investigators may terminate the study or report the results before all subjects realise their events. If no accidental loses or subject withdrawals, censored observations have times equal to the length of study time period; the censored time for each individual is the same and can be treated as a fixed time for a certain trial (Klein, 1997; Cox & Oakes, 1984; Kalbfleisch & Prentice, 2002).

Type II censoring

For Type II censoring, the study continues until r individuals experience the pre-specified event of interest. This number r may be some predetermined integer. Experiments for testing equipment failure time often involve this type of censoring. In this case, the censored time for each individual may be different and can be treated as a random variable. However, Type II censoring rarely occurs in clinical trials involving human subjects (Klein, 1997; Cox & Oakes, 1984; Kalbfleisch & Prentice, 2002).

Random censoring

Random censoring involves the so-called *competing risks* scenario. In this, individuals experience other competing events which may cause them to be removed from the study, and the primary event of interest is then not observed (p.61,Klein (1997)). For individuals in the HIP trial analysed in this thesis the underlying censoring mechanism maybe viewed as this type of *random censoring*.

3.1.4 Non-informative censoring

In general in survival analysis, it is often assumed that failure and censored times are independent. Hence we can assume that knowledge of a censoring time for an individual does not contribute further information about that individual's likelihood of survival at a future time. In other words, we can say that the distribution of censoring times contains no information about the distribution of the individual's failure time. Then the likelihood function with respect to failure time would not include any distribution for censoring time.

In other words, we have:

$$L(\boldsymbol{\theta}|T) \propto \prod_{i=1}^{n} f(t_i|\boldsymbol{\theta}_T)^{R_i} S(t_i|\boldsymbol{\theta}_T)^{1-R_i} \tag{3.1.3}$$

where $L(\boldsymbol{\theta}|T)$ denotes the likelihood function for survival time and $f(t_i|\boldsymbol{\theta}_T)$ refers to the probability density function at failure time t, while its associ-

ated survival function is denoted by $S(t_i|\boldsymbol{\theta}_T)$, and $\boldsymbol{\theta}_T$ represents the parameter associated with the failure time distribution. The likelihood form given in (3.1.3) highlights that the contribution from censored individuals is only through the survival function of failure time $S(t_i|\boldsymbol{\theta}_T)$. This is because an individual being censored at time t, $C_i = t$, is equivalent to that same individual surviving after time t ($T_i > t$). This is also known as non-informative censoring as no function of censoring time is involved in the likelihood structure with respect to failure time (equation (3.1.3)) (Klein, 1997; Kalbfleisch & Prentice, 2002).

3.1.5 Informative censoring

However, the estimation of the failure time distribution, such as $S(t)$ the survival function, may be influenced by the distribution of the censoring time. For instance, some individuals withdraw from a clinical trial because of the high risk of negative effects of receiving their treatment. Such cases are not treated as censored at random and censoring time distributions will influence the estimations of failure time distribution (Klein, 1997; Kalbfleisch & Prentice, 2002). Therefore, the distribution of censoring time needs to be included in the likelihood function.

Note that $(Y_i = t, R_i = 1)$ and $(T_i = t, C_i > t)$ (where Y_i and R_i denote the observed time and censoring indicator) indicate the same event, which is that individual i failed (experienced the event of interest) at time t. Consequently, the probability of occurrence of this event can be written as $Pr(Y_i = t, R_i = 1)$ and also $Pr(T_i = t, C_i > t)$. Both must be identical. Then for the *ith* individual experiencing the event of interest at time t, we have the following equation:

$$Pr(Y_i = t, R_i = 1; |X, \boldsymbol{\theta}) = Pr(T_i = t, C_i > t|X; \boldsymbol{\theta}) \tag{3.1.4}$$

where X denotes covariates vector and $\boldsymbol{\theta}$ is the parameter associated with the joint distribution of failure and censoring time.

Similarly, for the *ith* individual censored at time t, we have the equation

as follows:

$$Pr(Y_i = t, R_i = 0|X; \boldsymbol{\theta}) = Pr(T_i > t, C_i = t|X; \boldsymbol{\theta}) \tag{3.1.5}$$

where X is covariates vector and $\boldsymbol{\theta}$ represents the parameter associated with the joint distribution of failure and censoring time.

In general, we need additional assumptions to construct the likelihood function associated with the scenario of informative censoring as the likelihood form given in (3.1.3) is invalid for this situation. For the simplest case, when failure and censoring time are independent conditional on the covariates X, the likelihood function can be expressed as:

$$L(\boldsymbol{\theta}|T, C) \propto \prod_{i=1}^{n} [f(t_i|\boldsymbol{\theta}_T)G(t_i|\boldsymbol{\theta}_C)]^{R_i}[S(t_i|\boldsymbol{\theta}_T)e(t_i|\boldsymbol{\theta}_C)]^{1-R_i}, \tag{3.1.6}$$

where $L(\boldsymbol{\theta}|T, C)$ is the likelihood function for survival outcomes including both failure time and censoring time; $f(t_i|\boldsymbol{\theta}_T)$ and $S(t_i|\boldsymbol{\theta}_T)$ are same as defined in equation (3.1.3); $e(t)$ and $G(t)$ denote the probability density and probability functions at censoring time t.

Under the assumption of informative censoring, which is equivalent to that the failure and censoring times are independent, conditional on X, the joint probability of failure and censoring times given covariates X, $Pr(T_i, C_i; |X, \boldsymbol{\theta})$ can then be replaced by the product of the conditional probabilities of failure and censoring time, that is

$$\begin{aligned} Pr(T_i = t, C_i > t|X; \boldsymbol{\theta}) &= Pr(T_i = t|X; \boldsymbol{\theta}_T)Pr(C_i > t|X; \boldsymbol{\theta}_C) \quad \text{for the event,} \\ Pr(T_i > t, C_i = t|X; \boldsymbol{\theta}) &= Pr(T_i > t|X; \boldsymbol{\theta}_T)Pr(C_i = t|X; \boldsymbol{\theta}_C) \quad \text{for censoring.} \end{aligned} \tag{3.1.7}$$

In other words, standard likelihood structure formulating with respect to the failure time would not be suitable in such cases, because failure and censoring time are not independent exactly. Hence the parameter corresponding to failure time may also be determined by the censored time distribution.

3.2 Conventional Methodologies

Since the 1950s, many approaches, including the non-parametric, semi-parametric and parametric methods, have been developed for time-to-event studies. Most of the methods are based on the non-informative censoring assumption (Loeys & Goetghebeur, 2003; Robins & Tsiatis, 1991). In fact, to date, only a few papers exist on failure time analysis under the assumption of informative censoring, especially for latent ignorability (Frangakis & Rubin, 1999; Baker, 1998; Scharfstein et al., 2001) for censoring. The primary objective of many clinical trials is to provide a reliable comparison of the efficacy and safety of two interventions, where efficacy is often assessed in terms of a time-to-event outcome measure, such as the survival function or hazard function (Section 3.1.2).

3.2.1 Non-parametric methods

Non-parametric methods are one of the conventional statistical approaches for which there is no need to assume any distribution for the underlying data. These are also called parameter-free methods or distribution-free methods, which were developed to be used in cases where the researcher knows nothing about the parameters of the variable of interest. (Fraser, 1957; Kraft & Eeden, 1968; Casella & Berger, 2002).

In survival analysis, the most plausible non-parametric estimator of the survival function is the Kaplan-Meier (1958) or the Nelson-Aalen (1972,1978) estimator of $S(t)$ (Klein, 1997; Cox & Oakes, 1984; Kalbfleisch & Prentice, 2002; Ibrahim et al., 2001). Both of these are calculated under the assumption of non-informative censoring. Let t_1 represent the trial start time and d_i denotes the number of deaths in the *ith* time interval; the number of subjects under risk in the associated time interval is denoted by Y_i. Then the Kaplan-Meier estimator (p.84, Klein (1997)) of the survival function is formulated as:

$$\hat{S}(t) = \begin{cases} 1 & \text{if } t < t_1, \\ \prod_{t_i \leq t}(1 - \frac{d_i}{Y_i}) & \text{if } t_1 < t \ . \end{cases} \tag{3.2.1}$$

Nelson in 1972 proposed an alternative estimator of the cumulative hazard rate function in a reliability context; Aalen (1978) rediscovered the estimator by using modern counting process methodology (Aalen, 1975), so therefore, this estimator is also commonly known as the Nelson-Aalen estimator of the cumulative hazard

(Klein, 1997).

Let $\hat{H}(t)$ denote the estimated Nelson-Aalen estimator of the cumulative hazard function. It can then be written as follows:

$$\hat{H}(t) = \begin{cases} 0 & \text{if } t < t_1, \\ \sum_{t_i \leq t} \frac{d_i}{Y_i} & \text{if } t_1 < t. \end{cases} \tag{3.2.2}$$

For the relationship between the survival function and the cumulative hazard function, it is straightforward to obtain the estimator of survival function based on the Nelson-Aalen estimator which then has a form as follows:

$$\tilde{S}(t) = exp\left(-\hat{H}(t)\right). \tag{3.2.3}$$

Hypothesis test methods would also be an integral part of non-parametric methods for clinical scenarios. Of these, the most plausible and most popular test method is the so-called log-rank test method (p.193, Klein (1997)). Alternative test methods could be used, such as the Peto-Peto-Wilcoxon test (p.194, Klein (1997)) and the Fleming-Harrington test (p.194, Klein (1997)). We omit the context of these test methods as the non-parametric method is not the focus of the thesis.

3.2.2 Semi-parametric methods

In time-to-event studies, we often need to link the outcome of interest with so-called associated explanatory variables $\mathbf{X}$. This is out of the scope of the non-parametric method. Cox (1972), however, proposed a proportional hazards model, which has the following formulation:

$$h(t|\mathbf{X}) = h_0(t)\psi(\beta^T X); \tag{3.2.4}$$

where $h_0(t)$ is an arbitrary baseline hazard rate which is treated nonparametrically (p.230, Klein (1997)), $\psi(\beta^T X)$ represents a known function, and β is a parameter vector associate with $\mathbf{X}$, and ^{T} denotes transpose.

This model is known a semi-parametric model. Since $h(t|\mathbf{X})$ must be positive, a common form for ψ is the exponential function, $\psi(\beta^T X) = exp(\beta^T X)$. This model is also called the proportional hazard model because the hazard ratio is a constant for time-independent covariates $\mathbf{X}$. That is:

$$\frac{h(t|\mathbf{X_1})}{h(t|\mathbf{X_2})} = \frac{h_0(t)exp(\beta^T X_1)}{h_0(t)exp(\beta^T X_2)} = exp(\beta^T(X_1 - X_2); \tag{3.2.5}$$

where the model parameter β can be estimated by maximizing the related partial likelihood function.

Let $\mathcal{R}(t_i)$ be the set of all individuals who are still under study at a time just prior to t_i. Here t_i represents the *ith* ordered failure time ($t_1 < t_2 \cdots < t_I$). Then the partial likelihood function, assuming no ties in the event time data, has a form as follows:

$$L(\beta) = \prod_{i=1}^{I} \frac{exp(\beta^T X_i)}{\sum_{j \in \mathcal{R}(t_i)} exp(\beta^T X_j)}. \tag{3.2.6}$$

In the presence of ties among the event times, several suggestions for constructing the partial likelihood times have been given by Breslow (1974) and Efron (1977) (p.239, Klein (1997)).

3.2.3 Parametric methods

Parametric methods are based on the assumption that the outcome of interest follows some specific distribution. In survival analysis, failure time distributions have been considered for a so-called homogeneous population. In general, distributions such as the exponential, Weibull, log-normal, Gamma, and log-logistic distributions have been used throughout the failure time literature. The exponential and Weibull distributions have a closed-form expression for tail area probabilities, and also have simple formulas for the survival and hazards functions. Log-normal and Gamma distributions are also used frequently despite being generally less convenient for computation (Klein, 1997; Cox & Oakes, 1984; Kalbfleisch & Prentice, 2002).

It is common to estimate the parameters involved in the given distribution by a conventional statistical method, such as the maximum likelihood method (Casella & Berger, 2002). The estimated values of hazard function and survival function can then be easily obtained (Klein, 1997; Cox & Oakes, 1984; Kalbfleisch & Prentice, 2002).

3.3 Inference for causal effects for time-to-event data

As a very special variable, time has its own features, especially for time-to-event outcomes. For instance, a failure time has rarely a negative value and it never decreases. Therefore, distributions commonly used for general outcome or response variables, such as a normal distribution (in which negative values are often included) or a Bernoulli distribution (which is only suitable for binary variables)

do not suit specific data collected in time-to-event studies, where the outcome of interest is event time. Because event time is very unlikely to follow these previously mentioned distributions, the outcome of interest in the time-to-event studies should be distinguished from the general data scenario. This means that the event times should have their own distributions rather than conform to normal or Bernoulli distributions. In fact, many distributions are suitable for non-negative variables, such as the Weibull, log-normal, Gamma, etc. distributions. However, as noted in Chapter 2, potential-outcome models based on the *individual causal effect* (ICE) mostly focus on estimating the *average causal effect* (ACE) or *complier average causal effect* (CACE) for noncompliance for general data rather than time-to-event data. Our approach extends the ICE based model, as found in the study of O'Malley & Normand (2005) and Peng et al. (2004), for causal inference on survival data. Since in time-to-event studies, as discussed in Section 3.1.2, specific functions of failure time, e.g. the *Survival function* and the *hazard rate function*, are more interesting than failure time itself, it would be good progress if the causal effect on the time-to-event data can be expressed as a function of failure time. To reach this goal, we first need to define a new form of outcome of interest for time-to-event data.

3.3.1 A new outcome of interest

Let us consider the indicator function $I\{\}$, which has a value of 1 when the condition given in $\{\}$ holds and 0 otherwise. We then use v to denote the potential treatment level and denote $O_{iv}(t) = I\{T_i(v) > t\}$ as the outcome of interest, which is the indicator of individual i experiencing treatment level v and surviving after time t if assigned to treatment level v. Note that the new outcome $O_{iv}(t)$ is a function of time and also a new outcome variable.

3.3.2 Causal effect for time-to-event data

By following the general definition of *Individual Causal Effect*, *ICE* corresponding to the new outcome $O_{iv}(t)$ has the form as follows:

$$ICE_i(t) = O_{i1}(t) - O_{i0}(t) = I\{T_i(1) > t\} - I\{T_i(0) > t\}, \tag{3.3.1}$$

which is a function of failure time t.

3.3.3 Average causal effect (ACE) for time-to-event data

We can then define a time variant of ACE as follows:

$$\begin{aligned} ACE(t) &= E(O_{i1}(t) - O_{i0}(t)) = E(O_{i1}(t)) - E(O_{i0}(t)) \\ &= \frac{\sum_{i=1}^{\mathbb{N}} I\{T_i(1) > t\}}{\mathbb{N}} - \frac{\sum_{i=1}^{\mathbb{N}} I\{T_i(0) > t\}}{\mathbb{N}}, \end{aligned} \tag{3.3.2}$$

where $\mathbb{N}$ denotes the population size.

Let $\Gamma_v(t)$ denote the term $\frac{\sum_{i=1}^{\mathbb{N}} I\{T_i(v)>t\}}{\mathbb{N}}$ in equation (3.3.2), which is the proportion of individuals who would survive beyond time t if assigned to treatment level v. Then equation (3.3.2) can be expressed as

$$ACE(t) = \Gamma_1(t) - \Gamma_0(t). \tag{3.3.3}$$

Can we link the new function of time, $ACE(t)$, to the survival function ?

Recall that v represents the potential treatment level, the numerator of $\Gamma_v(t)$ is $\sum_{i=1}^{\mathbb{N}} I\{T_i(v) > t\}$, which gives the number of individuals in the study if assigned to treatment level v who survive after time t; while the denominator $\mathbb{N}$ is the total number of individuals in the population. $\Gamma_v(t)$ is then the survival proportion of individuals if assigned to treatment level v who survive beyond time t in the population. Now the probability of an event can be expressed as a proportion of the number of events and the total number of cases (population), that is:

$$\text{Probability of an event } = \frac{\text{Number of occurrence of event}}{\text{Population}}. \tag{3.3.4}$$

Thereby, the fraction $\Gamma_v(t)$ can be written as:

$$\Gamma_v(t) = \frac{\text{Number of event } \{T_i(v) > t\}}{\text{Population}}. \tag{3.3.5}$$

The numerator of $\Gamma_v(t)$ is the total number of occurrences of events of interest (survival after time t), and its denominator is the total number of individuals involved (the population).

Therefore, as stated before, the fraction $\Gamma_v(t)$ can be viewed as the probability of an individual surviving after time t with the potential treatment level v. Recall that the survival function $S(t) = Pr(T_i > t)$ is the probability of an individual surviving beyond time t. Here the probability notation Pr() can be interpreted as the population proportion or as a relative frequency under repeated sampling

from the population. For the potential treatment level v, $S_v(t)$ can be written as $S_v(t) = Pr(T(v) > t)$. Hence the two quantities $\Gamma_v(t)$ and $S_v(t)$ describe the same object, the survival probability of individuals exposed to potential level v. Therefore we have $\Gamma_v(t) \equiv S_v(t)$. After replacing $\Gamma_v(t)$ with $S_v(t)$ (for $v = 1$ or 0) in equation (3.3.3), we have obtained a relationship between $ACE(t)$ and $S_v(t)$ (for $v = 1$ 0r 0), i.e.:

$$ACE(t) = S_1(t) - S_0(t). \tag{3.3.6}$$

Equation (3.3.6) successfully extends the potential outcomes model for the general data (which are based on the *individual causal effect* rather than the *related causal effect*) to the time-to-event scenario, the new form of outcome $O_{iv}(t) = I\{T_i(v) > t\}$ provides a useful bridge to achieve this. A more important point is that $O_{iv}(t)$ has linked our new quantity of $ACE(t)$ in causal inference with the conventional quantity, $S_v(t)$, in time-to-event studies. It would be convenient to adopt traditional methods for evaluating the survival function in order to estimate the *average causal effect*. In addition, if we regard the survival function, $S_v(t)$, as special in time-to-event data analysis, as is the mean value μ_v in general data analysis, it should be straightforward to employ the conventional methods developed in the literature on ACE to the $ACE(t)$ extension.

The fundamental problem described and defined in Section 1.4.4 still exists, however, because for each individual i, only one of its outcomes of interest, $O_{i1}(t)$ or $O_{i0}(t)$, is observable, and likewise for their survival functions. Nevertheless, as we noted in the previous chapter, randomisation may serve as a powerful tool to handle this problem. Consider this conceptual link between $ACE(t)$ and $Z_i = 0$ or 1 for RCT settings. In RCTs, the survival function $S_v(t)$ for each possible treatment group is often estimated via those observations involved in the relevant treatment group. For instance, all individuals assigned to the treated group ($Z_i = 1$) would be a source of information to estimate $S_1(t)$, that is $\hat{S}_1(t)$=$S_{Z=1}(t)$. Similarly, for the control group, we have $\hat{S}_0(t)$=$S_{Z=0}(t)$. Consequently, the estimated *average causal effect* $\widehat{ACE}(t)$ can be obtained as follows:

$$\widehat{ACE}(t) = \hat{S}_1(t) - \hat{S}_0(t). \tag{3.3.7}$$

3.3.4 Complier average causal effect (CACE) for time-to-event data

The *Complier Average Causal Effect* (CACE) for time-to-event data is also a function of failure time, which we create via $CACE(t)$. Let $O_{i,uz}(t) = I\{T_i(z) >$

$t|U_i = u\}$ denote the outcome of interest, where $U_i = u$ is the compliance type introduced in Chapter 1. We then have

$$\widehat{CACE}(t) = \hat{S}_{c1}(t) - \hat{S}_{c0}(t), \tag{3.3.8}$$

where $\hat{S}_{cz}(t)$ represents the estimated survival function at time t corresponds to the subgroups where all individuals have compliance type $U_i = c$ and have been assigned to the treated group $Z = 1$ or to the control group $Z = 0$. That is:

$$\hat{S}_{cz}(t) = \frac{\sum_{i=1}^{N_z} I\{T_i(z) > t|U_i = c\}}{N_z}, \quad \text{for } z = 0 \text{ or } 1 \tag{3.3.9}$$

3.4 Summary

This chapter provides the new definitions of *Average Causal Effect* (ACE) and *Compliers Average Causal Effect* (CACE) for time to-event data, which allow us to extend those potential-outcome models based on the ICE (Individual Causal Effect) for general data to be suitable for time-to-event data in a straightforward manner. Note that, unlike general data, time-to-event data has its special features, among them *censoring* is a primary issue that needs to be considered carefully, which can be viewed as a special case of non-response.

In Section 3.3, we developed the ICE based potential-outcome model for time-to-event data through an indicator function $I\{T_i(v) > t\}$ and the newly defined variable $O_{iv}(t) = I\{T_i(v) > t\}$. The advantage of this extension is that it successfully links quantities of the *causal effect*, say ACE and $CACE$, with the associated survival function of failure time. These functions of time, namely $ACE(t)$ and $CACE(t)$, will reflect more features of the failure time distributions and also avoid effects from censoring on failure time. In addition, it is convenient to adopt conventional methods for evaluating survival function to estimate the ACE(t) and the CACE(t) by following our new definitions. As a consequence, more information can be obtained via our new ACE(t) or CACE(t) than from the conventional ACE or CACE, as the latter are not functions of failure time.

Furthermore, it would also be convenient to employ those methods that developed for causal inference on general data through estimating the ACE and the CACE in time-to-event studies. For instance, the method of Peng et al. (2004) which considered covariates in regression models could also be adopt for time-to-event data.

Chapter 4

The ignorable and non-ignorable censoring mechanisms

As stated in previous chapters (Chapter 1 and Chapter 3), *censoring* is the primary feature of time-to-event data (Sections 1.3 and 3.1). When censoring occurs, an individual's event time, the outcome of interest, cannot be observed. This is a special case of non-response or of a so-called missing outcome. This chapter is based on the theory of missing data (Little & Rubin, 2002) and defines three censoring mechanisms, which include censoring at the end of the follow-up (Section 4.3.1); censoring at random conditional on observed covariates only (Section 4.3.2); and censoring at random conditional on observed and unobservable covariates (Section 4.3.3). In addition, three likelihood forms corresponding to these three censoring mechanisms are constructed. Consequently, the ignorability of these three censoring mechanisms can be shown, which can be viewed as the foundation for creating our potential-outcome models. These models are given in subsequent chapters of this thesis. However, in order to simplify the presentation, in this chapter we consider only single outcome variables rather than potential-outcome vectors.

4.1 Missing data mechanisms

Little & Rubin (2002) pointed out that missing data mechanisms *per se* are crucial, as missing data methods strongly depend on the nature of the dependencies in the missing mechanism. However, in the analysis of data with missing values, the important role of the missing mechanism has been largely ignored. This continued until Rubin (1976) addressed the concept of missing data theoretically through missing data indicators. Rubin's missing data theory, in brief, claimed that the missing mechanism can be ignored only when it is missing completely at random (MCAR) or missing at random (MAR) (Little & Rubin, 2002).

For convenience, we use the same notation as in Little & Rubin (2002). Let $\mathbb{Y} = (y_{ij})$ denote the complete data, an $(n \times K)$ rectangular dataset with missing

values, with *ith* row $y_i = (y_{i1},\ y_{i2},\ \ldots\ y_{iK})$, where y_{ij} is the value of the *jth* variable $\mathbb{Y}_j$ for the *ith* subject. The missing data indicator matrix is denoted by $M = (m_{ij})$, i.e. $m_{ij} = 1$ if y_{ij} is missing and $m_{ij} = 0$ if y_{ij} is observed. Then the missing data mechanism is characterized by the conditional distribution of M given $\mathbb{Y}$, say $f(M|\mathbb{Y}, \phi)$, where ϕ is a set of unknown parameters.

4.1.1 Missing completely at random (MCAR)

For the case of data missing completely at random (MCAR), it is assumed that the "missingness" does not depend on the value of the data $\mathbb{Y} = (\mathbb{Y}_{obs}, \mathbb{Y}_{mis})$, including the missing outcomes $\mathbb{Y}_{mis}$ or of the observed $\mathbb{Y}_{obs}$. This can be expressed as follows:

$$f(M|\mathbb{Y}, \phi) = f(M|\phi) \quad \text{for all } \mathbb{Y} \text{ and } \phi. \tag{4.1.1}$$

4.1.2 Missing at random (MAR)

For the case of data missing at random (MAR), it is supposed that missingness only depends on the observed value of the data $\mathbb{Y}_{obs}$, instead of the missing value of the data $\mathbb{Y}_{mis}$, such that:

$$f(M|\mathbb{Y}, \phi) = f(M|\mathbb{Y}_{obs}, \phi) \quad \text{for all } \mathbb{Y}_{mis} \text{ and } \phi. \tag{4.1.2}$$

4.1.3 Not missing at random (NMAR)

If the distribution of M depends on the missing values in the data matrix $\mathbb{Y}$, the mechanism is called not missing at random (NMAR) (Little & Rubin, 2002). We then have:

$$f(M|\mathbb{Y}, \phi) = f(M|\mathbb{Y}_{obs}, \mathbb{Y}_{mis}, \phi) \quad \text{for } \phi. \tag{4.1.3}$$

4.2 Censoring mechanisms

It is necessary to consider the underlying potential censoring mechanism carefully in time-to-event studies, especially when formulating the likelihood structure, (as censoring can be viewed as a special type of missing data). In survival analysis, we often denote time-to-event data as a pair of (Y_i, R_i). Here the censoring indicator R (1 for the event occurring (observed), 0 for censoring (missing)) indicates the situation of observed or missing failure time. However, it is noteworthy that censoring time does provide information beyond the censoring indicator itself. For

instance, in the case of right censoring time $C_i = t$, we know that $T_i > t$ must hold, even though observing C_i precludes full observation of T_i.

To discuss the incomplete data issue in time-to-event studies, it is of value to consider instead the distribution of $(T_i,\ C_i)$ (the failure time T_i and the censoring time C_i) rather than the observed time and its censoring indicator $(Y_i,\ R_i)$. This is so, even despite the predominance in terms of the $(Y_i,\ R_i)$ formulation in conventional survival analysis.

Since censoring causes a special case of missing outcome where failure time is missing, it is important to ensure which type of censoring mechanism has occurred, because the likelihood construct may vary according to the censoring mechanism.

In the following subsections, three types of mechanism for censoring are defined, which derive from three types of missingness as defined in Section 4.1.

4.2.1 Censoring completely at random

In time-to-event studies, MCAR is analogous to censoring completely at random. In this case, the censoring time distribution does not depend on any observed or missing data. Thereby, we can write:

$$Pr(\mathbf{C}|\mathbf{T}, \mathbf{D}, \mathbf{Z}, \mathbf{U}, \mathbf{X}, \boldsymbol{\theta}_C) = Pr(\mathbf{C}|\boldsymbol{\theta}_C) \quad \text{for all} \quad \mathbf{T}, \mathbf{D}, \mathbf{Z}, \mathbf{U}, \mathbf{X} \,\text{and}\, \boldsymbol{\theta}_C. \quad (4.2.1)$$

Recall that in this chapter we only focus on single variable rather than potential outcomes vector, which means that $\mathbf{T}$ and $\mathbf{C}$ are observable failure and censoring time; $\mathbf{Z}$ and $\mathbf{D}$ denote the randomised assigned and actually received treatment; $\mathbf{U}$ is unobservable compliance type, and $\mathbf{X}$ is covariate vector.

4.2.2 Censoring at random

In time-to-event studies, MAR is analogous to censoring at random, which means that the censoring time distribution depends only on the observed covariates. They are, in the case of a RCT with noncompliance, the assigned treatment $Z_i = z$, the actual received treatment $D_i = D_i^{obs}$ and other pretreatment covariates X. In this case, we suppose that the censoring mechanism depends only on these observable covariates $\mathbf{Z}$, $\mathbf{D}(Z)$ and $\mathbf{X}$, we can then write:

$$Pr(\mathbf{C}|\mathbf{T}, \mathbf{D}, \mathbf{Z}, \mathbf{U}, \mathbf{X}, \boldsymbol{\theta}_C) = Pr(\mathbf{C}|\mathbf{D}, \mathbf{Z}, \mathbf{X}, \boldsymbol{\theta}_C) \quad \text{for all } \mathbf{T},\ \mathbf{U} \text{ and } \boldsymbol{\theta}_C. \quad (4.2.2)$$

4.2.3 Censoring not at random

Censoring not at random implies that the censoring time distribution may not only depend on observed data, but also on the missing data. For RCTs with noncompliance, we suppose that the censoring time distribution depends on the observable covariates, (D_i, Z_i and X_i), and also on the unobservable covariate U_i. This assumption can be expressed as follows:

$$Pr(\mathbf{C}|\mathbf{T},\mathbf{D},\mathbf{Z},\mathbf{U},\mathbf{X},\boldsymbol{\theta}_C) = Pr(\mathbf{C}|\mathbf{D},\mathbf{U},\mathbf{Z},\mathbf{X},\theta_C) \quad \text{for all } T \text{ and } \boldsymbol{\theta}_C. \qquad (4.2.3)$$

Specifically in this thesis, we focus on studying a special case, in which the censoring time ($\mathbf{C}$) and failure time ($\mathbf{T}$) are conditionally independent given the unobservable covariates ($\mathbf{U}$). This may be viewed as the assumption of *latent ignorability* for time-to-event data (Frangakis & Rubin, 1999). In brief, it can be expressed as:

$$Pr(\mathbf{T},\mathbf{C}|\mathbf{U},\mathbf{X},\mathbf{Z}) = Pr(\mathbf{T}|\mathbf{U},\mathbf{X},\mathbf{Z})Pr(\mathbf{C}|\mathbf{U},\mathbf{X},\mathbf{Z}) \qquad (4.2.4)$$

(Frangakis & Rubin, 1999; Mealli et al., 2004).

4.3 Likelihood structures

Having defined the three censoring mechanisms, we can then establish the relevant likelihood structure. At first, we consider the joint distribution of all related variables, including failure (T_i) and censoring (C_i) times, assigned treatment (Z_i), received treatment ($D_i(Z_i)$), compliance type (U_i) and other pre-treatment covariates (X_i). Note that in this chapter, each outcome variable is viewed as a single outcome rather than a potential outcomes vector. So the failure (T_i) and censoring (C_i) time indicate single outcomes only, which correspond to the treatment levels $Z_i = z$, i.e. $T_i = T_i(z)$ and $C_i = C_i(z)$. For the received treatment (D_i), only the observed $D_i = D_i(z)$ is considered.

Suppose N individuals are included in the RCT. Then the assigned treatment is a vector of length N, $\mathbf{Z} = (Z_1, \ldots, Z_N)^T$. Consequently, the single outcomes can also be viewed as a set of vectors of length N, where each component is a scalar. We have:

> $\mathbf{D} = (D_1, \ldots, D_N)^T$: the actually received treatment corresponding to treatment group $Z_i = z$;

$\mathbf{T} = (T_1, \ldots, T_N)^T$: the failure time associated with the given treatment $Z_i = z$; and

$\mathbf{C} = (C_1, \ldots, C_N)^T$: the censoring time associated with the given treatment $Z_i = z$.

Similarly, covariate variables can still be written in vector forms of length N, as follows: $\mathbf{U} = (U_1, \ldots, U_N)^T$ and $\mathbf{X} = (X_1, \ldots, X_N)^T$. Of the vector variables for the N individuals given above, only the component of pre-treatment X_i is allowed to be a vector when the number of pre-treatment covariates (denoted by J) involved in the RCT is more than one ($J > 1$), that is $X_i = (X_{i1}, \ldots, X_{iJ})$.

By the *latent ignorability* assumption (Frangakis & Rubin, 1999), the joint distribution of failure ($\mathbf{T}$) and censoring ($\mathbf{C}$) times that are conditional on all observable ($\mathbf{Z}$, $\mathbf{D}$, $\mathbf{X}$) and on unobservable covariates ($\mathbf{U}$) is denoted by $Pr(\mathbf{T}, \mathbf{C}|\mathbf{D}, \mathbf{Z}, \mathbf{X}; \boldsymbol{\theta})$, and can be written as:

$$Pr(\mathbf{T}, \mathbf{C}|\mathbf{D}, \mathbf{Z}, \mathbf{X}, \mathbf{U}; \boldsymbol{\theta}) = Pr(\mathbf{T}|D, \mathbf{Z}, X, \mathbf{U}, \boldsymbol{\theta}_T) Pr(\mathbf{C}|D, Z, \mathbf{X}, \mathbf{U}; \boldsymbol{\theta}_C), \quad (4.3.1)$$

where $\boldsymbol{\theta}_T$ and $\boldsymbol{\theta}_C$ denote the parameters corresponding to the conditional probabilities of failure time and censoring time, respectively. In fact, equation (4.3.1) can be simplified further when the censoring mechanism is specified as one of the three censoring mechanisms discussed in Section 4.2.

4.3.1 Censoring at the end of follow-up

In this situation, the censored individuals, namely subjects who did not experience the event of interest (during the observation period), are all censored at the same time (on the same day) when the trial ends, given that they all enrolled into the trial at the same time (on the same day), or given that they have been followed up for the same time duration. Therefore their censoring times are the same; i.e. the censoring time C_i (for $i \in \{1,\ 2, \ldots\ N_C\}$; where N_C is the number of censored individuals in the RCT) is considered to be as a constant, $C_i = Cr$, rather than a random variable. Therefore, in this case, no censoring distribution or censoring mechanism exists. This implies that a probabilistic model for censoring time is not required.

Consequently, in such a case, instead of the joint distribution of failure and censoring time, $Pr(\mathbf{T}, \mathbf{C}|\mathbf{D}, \mathbf{Z}, \mathbf{X}, \mathbf{U}; \boldsymbol{\theta})$, only the conditional distribution of failure time $\mathbf{T}$ (given those observable and unobservable covariates) is involved. Therefore

equation (4.3.1) is simplified as follows:

$$Pr(\mathbf{T}|\mathbf{D},\mathbf{Z},\mathbf{X},\mathbf{U};\boldsymbol{\theta}) = Pr(\mathbf{T}|\mathbf{D},\mathbf{Z},\mathbf{X},\mathbf{U};\boldsymbol{\theta}_T), \tag{4.3.2}$$

where the censoring mechanism is ignored because the so-called censoring time C_i is a constant $C_i = Cr$ (for $i \in \{1,\ 2, \dots\ N_C\}$).

Therefore, the parameter $\boldsymbol{\theta}$ associated with the joint distribution of failure and censoring time is only associated with the parametric of the failure time distribution, $\boldsymbol{\theta}_T$. To estimate the parameter $\boldsymbol{\theta}_T$, we need only to construct the likelihood function with respect to the parameter $\boldsymbol{\theta}_T$, which then only relates to the distribution of failure time t, $Pr(\mathbf{T}|D,\mathbf{Z},\mathbf{X},\mathbf{U};\boldsymbol{\theta}_T)$.

Consider the joint distribution of failure time $\mathbf{T}$ and all the covariates (including the observable $\mathbf{D}$, $\mathbf{Z}$, $\mathbf{X}$ and unobservable $\mathbf{U}$). By following the relationship between the outcome and covariates, for the *ith* individual, this joint distribution can be written as the product of the conditional distribution of the outcome of interest (failure time T_i) and the joint distribution of all related covariates, namely the assigned treatment Z_i, the received treatment $D_i = D_i(Z_i)$, and the observed covariates X_i. In other words:

$$Pr(T_i, D_i, Z_i, X_i, U_i; \boldsymbol{\Psi}) = Pr(T_i|D_i, Z_i, X_i, U_i; \boldsymbol{\theta}) Pr(D_i, U_i, Z_i, X_i; \Upsilon), \tag{4.3.3}$$

where $\boldsymbol{\Psi}$ denotes the parameter associated with the joint distribution, which has been divided into two parts: $\boldsymbol{\Psi} = (\boldsymbol{\theta}, \Upsilon)$. One is the parameter $\boldsymbol{\theta}$ that corresponds to the conditional probability of the outcome variable (failure time T_i), given all related covariate variables (including the unobservable covariates U_i), and on all those observed covariates (the assigned treatment Z_i, the received treatment $D_i = D_i(Z_i)$ and the observed covariates X_i). The other parameter Υ corresponds to the joint distribution of all the covariate variables.

In addition, by following the property of probability, the second term on the right-hand side of equation (4.3.3), $Pr(D_i, U_i, Z_i, X_i|\Upsilon)$, can be broken down further as follows:

$$Pr(D_i, U_i, Z_i, X_i; \Upsilon) = Pr(U_i|D_i, Z_i, X_i; \boldsymbol{\pi}) Pr(D_i, Z_i, X_i; \lambda). \tag{4.3.4}$$

where $Pr(U_i|D_i, Z_i, X_i; \boldsymbol{\pi})$ (the right-hand side of equation (4.3.4)) denotes the conditional distribution of compliance type $U_i = u$, given all observed covariates $D_i,\ Z_i,\ X_i$, and $\boldsymbol{\pi}$ represents its associated parameter. $Pr(D_i, Z_i, X_i; \lambda)$ is the joint

distribution of all observed covariates D_i, Z_i, X_i, and λ is its associated parameter.

Equation (4.3.3) can then be re-written as:

$$\begin{aligned} &Pr(T_i, D_i, Z_i, X_i, U_i; \boldsymbol{\Psi}) \\ &\quad = Pr(T_i|U_i, D_i, Z_i, X_i; \boldsymbol{\theta}) Pr(U_i|D_i, Z_i, X_i; \boldsymbol{\pi}) Pr(D_i, Z_i, X_i; \lambda). \end{aligned} \tag{4.3.5}$$

In general, the likelihood function needs to be determined by the probability density function (p.d.f.) or cumulative probability function (c.d.f.) of the related outcome variable (given all observed data). For the parameter $\boldsymbol{\theta}$, given that U_i is theoretically unobservable, the likelihood function cannot be obtained directly from $Pr(T_i|U_i, D_i, Z_i, X_i; \boldsymbol{\theta})$, the so-called conditional distribution of failure time (T_i), given all covariates.

We then consider the joint distribution of failure time and all related covariates $Pr(T_i, D_i, Z_i, X_i, U_i; \boldsymbol{\Psi})$. By taking the integral of $Pr(T_i, D_i, Z_i, X_i, U_i; \boldsymbol{\Psi})$ with respect to $U_i^{mis} = U_i$, the joint distribution of failure time and all observed covariates $Pr(T_i, D_i, Z_i, X_i; \boldsymbol{\Psi})$ is obtained, that is:

$$\begin{aligned} &Pr(T_i, D_i, Z_i, X_i; \boldsymbol{\Psi}) = \\ &\int Pr(T_i|D_i, Z_i, X_i, U_i; \boldsymbol{\theta}) Pr(D_i, Z_i, X_i; \lambda) Pr(U_i|D_i, Z_i, X_i; \boldsymbol{\pi}) dU_i. \end{aligned} \tag{4.3.6}$$

Because of the assumption of independent and identically distributed (i.i.d.) observations (for all individuals), the joint distribution of failure time and covariates for all individuals, $Pr(\mathbf{T}, \mathbf{D}, \mathbf{Z}, \mathbf{X}, \mathbf{U}; \boldsymbol{\Psi})$, equals the multiple of the same joint distribution in each individual $Pr(T_i, D_i, Z_i, X_i; \boldsymbol{\Psi})$, that is:

$$Pr(\mathbf{T}, \mathbf{D}, \mathbf{Z}, \mathbf{X}, \mathbf{U}; \boldsymbol{\Psi}) = \prod_{i=1}^{N} Pr(T_i, D_i, Z_i, X_i; \boldsymbol{\Psi}). \tag{4.3.7}$$

After substituting equation (4.3.6) into equation (4.3.7), we then have the form as follows:

$$\begin{aligned} &Pr(\mathbf{T}, \mathbf{D}, \mathbf{Z}, \mathbf{X}, \mathbf{U}; \boldsymbol{\Psi}) \\ &\quad = \prod_{i=1}^{N} \int Pr(T_i|D_i, Z_i, X_i, U_i; \boldsymbol{\theta}) Pr(D_i, Z_i, X_i; \lambda) Pr(U_i|D_i, Z_i, X_i; \boldsymbol{\pi}) dU_i \\ &\quad = \prod_{i=1}^{N} Pr(D_i, Z_i, X_i; \lambda) \int Pr(T_i|D_i, Z_i, X_i, U_i; \boldsymbol{\theta}) Pr(U_i|D_i, Z_i, X_i; \boldsymbol{\pi}) dU_i, \end{aligned} \tag{4.3.8}$$

where the second equation holds because $Pr(D_i, Z_i, X_i; \lambda)$ can be regarded as a constant with respect to U_i, as the variable of integration U_i is not involved in $Pr(D_i, Z_i, X_i; \lambda)$. Now $Pr(T_i|D_i, Z_i, X_i, U_i; \boldsymbol{\theta})$ can be specified by the event of interest or by censoring, that is:

$$Pr(T_i|D_i, Z_i, X_i, U_i; \boldsymbol{\theta}) = \begin{cases} Pr(T_i = t|D_i, Z_i, X_i, U_i; \boldsymbol{\theta}) = f(t|D_i, Z_i, X_i, U_i; \boldsymbol{\theta}) & R_i = 1,\ T_i = t, \\ Pr(T_i > C_r|D_i, Z_i, X_i, U_i; \boldsymbol{\theta}) = S(C_r|D_i, Z_i, X_i, U_i; \boldsymbol{\theta}) & R_i = 0, C_i = C_r\ , \end{cases} \tag{4.3.9}$$

where $R_i = 1$ represents the event occurring, (0 for censoring); $f()$ and $S()$ denote the probability density function (p.d.f.) and the survival function, respectively.

The contribution to the likelihood for an individual dying before the end of follow-up at time t is then :

$$Pr(T_i = t|D_i, Z_i, X_i, U_i; \boldsymbol{\theta}_T), \tag{4.3.10}$$

and the contribution to the likelihood of an individual censored at time $t = Cr$ is:

$$Pr(T_i > Cr|D_i, Z_i, X_i; \boldsymbol{\theta}_T). \tag{4.3.11}$$

Note that C_r is a constant and denotes the censoring time for all censored individuals.

The likelihood function with respect to the parameter θ, which is associated with the conditional distribution of failure time T_i given D_i, Z_i, X_i, U_i, needs to also relate to the parameter π that corresponds to the conditional probability of compliance type U_i given D_i, Z_i, X_i. We denote the likelihood function as $L_1(\boldsymbol{\theta}, \boldsymbol{\pi})$, where

$$L_1(\boldsymbol{\theta}, \boldsymbol{\pi}) = \int Pr(T_i|D_i, Z_i, X_i, U_i; \boldsymbol{\theta}) Pr(U_i|D_i, Z_i, X_i; \boldsymbol{\pi}) dU_i. \tag{4.3.12}$$

Equation (4.3.12) can then be re-written as follows:

$$
\begin{aligned}
& L_1(\boldsymbol{\theta}, \boldsymbol{\pi}) \\
& = \int \left\{ \begin{array}{l} Pr(T_i = t | D_i, Z_i, X_i, U_i; \boldsymbol{\theta})^{R_i} \\ \times Pr(T_i > Cr | D_i, Z_i, X_i, U_i; \boldsymbol{\theta})^{1-R_i} \\ \times Pr(U_i | D_i, Z_i, X_i; \pi) \end{array} \right\} dU_i; \\
& = \int \left\{ \begin{array}{l} f(t | D_i, Z_i, X_i, U_i; \boldsymbol{\theta})^{R_i} \\ \times S(Cr | D_i, Z_i, X_i, U_i; \boldsymbol{\theta})^{1-R_i} \\ \times Pr(U_i | D_i, Z_i, X_i; \boldsymbol{\pi}) \end{array} \right\} dU_i.
\end{aligned}
\tag{4.3.13}
$$

Equation (4.3.13) then provides the required likelihood structure for the special case of RCTs with noncompliance, where all censoring occurs at the end of follow-up $t = Cr$.

4.3.2 Censoring at random conditional on observed variables only

In practice, it is not common to have a constant censoring time for each censored individual, especially in the medical area. In this section, we discuss the more usual case, where censoring time is treated as a random variable rather than a constant.

We assume that the censoring time distribution is dependent only on observable covariates, its related formulation is given in equation (4.2.2). Then the joint distribution can be decomposed further, such that

$$
\begin{aligned}
& Pr(T_i, C_i, D_i, Z_i, X_i; \boldsymbol{\theta}, \lambda) \\
& \quad = Pr(T_i | D_i, Z_i, X_i; \boldsymbol{\theta}_T) Pr(C_i | D_i, Z_i, X_i; \boldsymbol{\theta}_C) Pr(D_i, Z_i, X_i; \lambda).
\end{aligned}
\tag{4.3.14}
$$

Note that in this case, compliance type U_i is not involved in equation (4.3.14) as it is not an observed covariate. Censoring time distribution is included, and the parameter $\boldsymbol{\theta}$ associated with the joint distribution of the failure and censoring time can then be broken down into two parts, $\boldsymbol{\theta} = (\boldsymbol{\theta}_T, \boldsymbol{\theta}_C)$, where $\boldsymbol{\theta}_T$ and $\boldsymbol{\theta}_C$ are the parameters corresponding to the conditional distribution of the failure and censoring time, respectively.

Focussing on the likelihood structure of $L(\boldsymbol{\theta})$, we note that the censoring distribution has to be included, as $\boldsymbol{\theta} = (\boldsymbol{\theta}_T, \boldsymbol{\theta}_C)$. Under this conditional random censoring model, the joint distribution of failure and censoring time can thus be

broken down as follows:

$$Pr(T_i, C_i|D_i, Z_i, X_i; \boldsymbol{\theta}) = Pr(T_i|D_i, Z_i, X_i; \boldsymbol{\theta}_T)Pr(C_i|D_i, Z_i, X_i; \boldsymbol{\theta}_C). \quad (4.3.15)$$

According to the definition of likelihood function, the likelihood function, denoted here by $L_2(\boldsymbol{\theta})$, should be a function of the parameter $\boldsymbol{\theta}$ and also be proportional to $Pr(T_i, C_i|D_i, Z_i, X_i; \boldsymbol{\theta})$. The latter is the joint distribution of failure and censoring time, which is conditional on the observed covariates D_i, Z_i, X_i. The contribution to the likelihood of an individual dying before the end of follow-up at time t, is then:

$$\begin{aligned} &Pr(Y_i = t, R_i = 1|D_i, Z_i, X_i; \boldsymbol{\theta}) \\ &\quad = Pr(T_i = t|D_i, Z_i, X_i; \boldsymbol{\theta}_T)Pr(C_i > t|D_i, Z_i, X_i; \boldsymbol{\theta}_C). \end{aligned} \quad (4.3.16)$$

On the other hand, the contribution to the likelihood of an individual censored at time t, is:

$$Pr(Y_i = t, R_i = 0|D_i, Z_i, X_i; \boldsymbol{\theta}) \quad = Pr(T_i > t|D_i, Z_i, X_i; \boldsymbol{\theta}_T)Pr(C_i = t|D_i, Z_i, X_i; \boldsymbol{\theta}_C). \quad (4.3.17)$$

Equations (4.3.16) and (4.3.17) are based on the fact that $(Y_i = t, R_i = 1)$ and $(T_i = t, C_i > t)$ represent the same scenario, i.e. individual i experienced the event of interest at time t. Therefore $Pr(Y_i = t, R_i = 1|D_i, Z_i, X_i; \boldsymbol{\theta})$ and $Pr(T_i = t, C_i > t|D_i, Z_i, X_i; \boldsymbol{\theta})$ both represent the conditional probability that individual i failed at time t.

In this case, the likelihood function of $\boldsymbol{\theta}$ can be represented as

$$\begin{aligned} L_2(\boldsymbol{\theta}) &= \prod_i \begin{bmatrix} (Pr(T_i = t|D_i, Z_i, X_i; \boldsymbol{\theta}_T)Pr(C_i > t|D_i, Z_i, X_i; \boldsymbol{\theta}_C))^{R_i} \times \\ (Pr(T_i > t|D_i, Z_i, X_i; \boldsymbol{\theta}_T)Pr(C_i = t|D_i, Z_i, X_i; \boldsymbol{\theta}_C))^{1-R_i} \end{bmatrix} \\ &= \prod_i \begin{bmatrix} Pr(T_i = t|D_i, Z_i, X_i; \boldsymbol{\theta}_T)^{R_i}Pr(C_i > t|D_i, Z_i, X_i; \boldsymbol{\theta}_C)^{R_i} \times \\ Pr(T_i > t|D_i, Z_i, X_i; \boldsymbol{\theta}_T)^{1-R_i}Pr(C_i = t|D_i, Z_i, X_i; \boldsymbol{\theta}_C)^{1-R_i} \end{bmatrix} \\ &= \begin{bmatrix} \prod_i Pr(T_i = t|D_i, Z_i, X_i; \boldsymbol{\theta}_T)^{R_i}Pr(T_i > t|D_i, Z_i, X_i; \boldsymbol{\theta}_T)^{1-R_i} \times \\ \prod_i Pr(C_i > t|D_i, Z_i, X_i; \boldsymbol{\theta}_C)^{R_i}Pr(C_i = t|D_i, Z_i, X_i; \boldsymbol{\theta}_C)^{1-R_i}\} \end{bmatrix} \\ &= L_2(\boldsymbol{\theta}_T)L_2(\boldsymbol{\theta}_C). \end{aligned} \quad (4.3.18)$$

Essentially, the equivalence $L_2(\boldsymbol{\theta}) = L_2(\boldsymbol{\theta}_T)L_2(\boldsymbol{\theta}_C)$ allows us to ignore the distribution of $\boldsymbol{\theta}_C$. This is true if we are interested only on $\boldsymbol{\theta}_T$, which parameterizes survival time. Therefore, the likelihood function associated with the failure time

parameter $\boldsymbol{\theta}_T$ would have a form as follows:

$$L_2(\boldsymbol{\theta}_T) = \prod_i Pr(T_i = t|D_i, Z_i, X_i; \boldsymbol{\theta}_T)^{R_i} Pr(T_i > t|D_i, Z_i, X_i; \boldsymbol{\theta}_T)^{1-R_i}. \quad (4.3.19)$$

Using the same notation as given in the previous section (Section 4.3.1), $L_2(\theta_T)$ can be expressed as the following:

$$L_2(\boldsymbol{\theta}_T) = \prod_{i=1}^{N} f(t|D_i, Z_i, X_i; \boldsymbol{\theta}_T)^{R_i} S(t|D_i, Z_i, X_i; \boldsymbol{\theta}_T)^{1-R_i}. \quad (4.3.20)$$

This demonstrates that when the censoring mechanism is at random only conditional on the observed covariates (MAR), the parameter $\boldsymbol{\theta}_C$ associated with the censoring time distribution is ignorable.

4.3.3 Censoring at random conditional on observed and unobserved variables

When censoring at random is conditional on unobserved variables, the process governing censoring is not considered independent of the process that governs survival. This holds even after conditioning on all observed covariates. However, the remaining dependency between censoring and survival is assumed to be accounted for by an unobserved covariate, U, the compliance type, (see definition given in equation (1.7.2) Section 1.7.2). Here, the censoring mechanism is considered to be at random, and conditional on either the observed covariates (D_i, Z_i, X_i) and on the unobserved covariate U_i. We then have the NMAR case where the censoring mechanism is not at random. Then equation (4.2.3) holds and the joint distribution can be written as follows:

$$\begin{aligned} &Pr(T_i, C_i, D_i, Z_i, U_i, X_i; \boldsymbol{\Psi}) \\ &= \left\{ \begin{array}{l} Pr(T_i|D_i, Z_i, U_i, X_i; \boldsymbol{\theta}_T) Pr(C_i|D_i, Z_i, U_i, X_i; \boldsymbol{\theta}_C) \\ \times Pr(U_i|D_i, Z_i, X_i; \boldsymbol{\pi}) Pr(D_i, Z_i, X_i; \lambda) \end{array} \right\} \end{aligned} \quad (4.3.21)$$

Recall that compliance type U_i is considered as an unobservable covariate. Then the joint distribution of all observed variables can then be obtained by integrating $Pr(T_i, C_i, D_i, Z_i, U_i, X_i; \boldsymbol{\psi})$ with respect to U_i, which has the form as

follows:

$$\begin{aligned} & Pr(T_i, C_i, D_i, Z_i, X_i; \boldsymbol{\psi}) \\ & = \int \left\{ \begin{array}{l} Pr(T_i|D_i, Z_i, U_i, X_i; \boldsymbol{\theta}_T) Pr(C_i|D_i, Z_i, U_i, X_i; \boldsymbol{\theta}_C) \\ \times Pr(U_i|D_i, Z_i, X_i; \boldsymbol{\pi}) Pr(D_i, Z_i, X_i; \lambda) \end{array} \right\} dU_i. \end{aligned} \tag{4.3.22}$$

Under random censoring that is conditional on unobservable covariate U_i, the joint distribution of failure and censoring times can be broken down as follows:

$$\begin{aligned} & Pr(T_i, C_i|D_i, Z_i, X_i; \boldsymbol{\theta}, \boldsymbol{\pi}) \\ & = \int Pr(T_i|D_i, Z_i, U_i, X_i; \boldsymbol{\theta}_T) Pr(C_i|D_i, Z_i, U_i, X_i; \boldsymbol{\theta}_C) Pr(U_i|D_i, Z_i, X_i; \boldsymbol{\pi}) dU_i. \end{aligned} \tag{4.3.23}$$

Now the contribution to the likelihood of an individual dying before the end of follow-up at time t is given by:

$$\begin{aligned} & Pr(Y_i = t, R_i = 1|D_i, Z_i, X_i; \boldsymbol{\theta}) \\ & = \int \left\{ \begin{array}{l} Pr(T_i = t|D_i, U_i, Z_i, X_i; \boldsymbol{\theta}_T) \\ \times Pr(C_i > t|D_i, U_i, Z_i, X_i; \boldsymbol{\theta}_C) \\ \times Pr(U_i|D_i, Z_i, X_i; \boldsymbol{\pi}) \end{array} \right\} dU_i. \end{aligned} \tag{4.3.24}$$

Meanwhile, the contribution to the likelihood of an individual censored at time t is as follows:

$$\begin{aligned} & Pr(Y_i = t, R_i = 0|D_i, Z_i, X_i; \boldsymbol{\theta}) \\ & = \int \left\{ \begin{array}{l} Pr(T_i > t|D_i, U_i, Z_i, X_i; \boldsymbol{\theta}_T) \\ \times Pr(C_i = t|D_i, U_i, Z_i, X_i; \boldsymbol{\theta}_C) \\ \times Pr(U_i|D_i, Z_i, X_i; \boldsymbol{\pi}) \end{array} \right\} dU_i. \end{aligned} \tag{4.3.25}$$

As noted earlier, the parameter $\boldsymbol{\pi}$, which is associated with the conditional distribution of compliance type U_i, has to be involved in the joint distribution. In this case, the two parameters, say $\boldsymbol{\theta} = (\boldsymbol{\theta}_T, \boldsymbol{\theta}_C)$ and $\boldsymbol{\pi}$, have to be included in the given likelihood function. This is the case, even though we are only interested in the parameter associated with the failure time distribution, namely $\boldsymbol{\theta}_T$.

$$
\begin{aligned}
& L_3(\boldsymbol{\theta}, \boldsymbol{\pi}) = \\
&= \prod_i \left\{ \begin{array}{l} \left[\int Pr(T_i = t|D_i, U_i, Z_i, X_i; \boldsymbol{\theta}_T) Pr(C_i > t|D_i, U_i, Z_i, X_i; \boldsymbol{\theta}_C) Pr(U_i|D_i, Z_i, X_i; \boldsymbol{\pi}) dU_i\right]^{R_i} \times \\ \left[\int Pr(T_i > t|D_i, U_i, Z_i, X_i; \boldsymbol{\theta}_T) Pr(C_i = t|D_i, U_i, Z_i, X_i; \boldsymbol{\theta}_C) Pr(U_i|D_i, Z_i, X_i; \boldsymbol{\pi}) dU_i\right]^{1-R_i} \end{array} \right\} \\
&= \prod_i \int \left\{ \begin{array}{l} \left[Pr(T_i = t|D_i, U_i, Z_i, X_i; \boldsymbol{\theta}_T) Pr(C_i > t|D_i, U_i, Z_i, X_i; \boldsymbol{\theta}_C)\right]^{R_i} \times \\ \left[Pr(T_i > t|D_i, U_i, Z_i, X_i; \boldsymbol{\theta}_T) Pr(C_i = t|D_i, U_i, Z_i, X_i; \boldsymbol{\theta}_C)\right]^{1-R_i} \end{array} \right\} Pr(U_i|D_i, Z_i, X_i; \boldsymbol{\pi}) dU_i \\
&= \prod_i \int \left\{ \begin{array}{l} Pr(T_i = t|D_i, Z_i, X_i; \boldsymbol{\theta}_T)^{R_i} Pr(T_i > t|D_i, U_i, Z_i, X_i; \boldsymbol{\theta}_T)^{1-R_i} \times \\ Pr(C_i > t|D_i, U_i, Z_i, X_i; \boldsymbol{\theta}_C)^{R_i} Pr(C_i = t|D_i, U_i, Z_i, X_i; \boldsymbol{\theta}_C)^{1-R_i} \end{array} \right\} Pr(U_i|D_i, Z_i, X_i; \boldsymbol{\pi}) dU_i.
\end{aligned}
\tag{4.3.26}
$$

We denote the likelihood function for this case as $L_3(\boldsymbol{\theta}, \boldsymbol{\pi})$, which has the form given in equation (4.3.26). Because both distributions of $\boldsymbol{\theta}_T$ and $\boldsymbol{\theta}_C$ are related to the parameter $\boldsymbol{\pi}$, which corresponds to the distribution of U_i (the unobservable covariate). This censoring case, unlike the case considered previously (Section 4.3.2), the likelihood function therefore cannot be broken down as $L_3(\boldsymbol{\theta}) \neq L_3(\boldsymbol{\theta}_T)L_3(\boldsymbol{\theta}_C)$. In this case, the censoring mechanism is considered to be non-ignorable. We must then construct the likelihood function of $\boldsymbol{\theta}$ rather than of $\boldsymbol{\theta}_T$ itself.

Equation (4.3.26) can then be expressed as follows

$$L_3(\boldsymbol{\theta}, \boldsymbol{\pi}) = \\ = \prod_i \int \left[\begin{array}{c} f(t|D_i, Z_i, X_i; \boldsymbol{\theta}_T)^{R_i} S(t|D_i, Z_i, X_i; \boldsymbol{\theta}_T)^{1-R_i} \times \\ g(t|D_i, Z_i, X_i; \boldsymbol{\theta}_C)^{1-R_i} G(t|D_i, Z_i, X_i; \boldsymbol{\theta}_C)^{R_i} \end{array} \right] pr(U_i|D_i, Z_i, X_i; \boldsymbol{\pi}) dU_i. \tag{4.3.27}$$

where $f()$ and $S()$ denote the probability density and survival function of the failure time; meanwhile, $g()$ and $G()$ denote these same name functions for censoring time.

4.4 Summary

We addressed three types of censoring mechanisms (Section 4.2). The first two can be treated as ignorable censoring mechanisms, whereas the last can be treated as a non-ignorable censoring mechanism.

Because of the presence of unobservable covariates, (i.e. compliance type) the likelihood function is much more complicated than in the case of an ignorable censoring mechanism. For situation, we provide two likelihood structures, equations (4.3.6) and (4.3.22) for time-to-event studies with noncompliance. Equation (4.3.6) can be viewed as a parametric survival model allowing noncompliance with an ignorable censoring mechanism; it is comparable to the approach proposed by Loeys & Goetghebeur (2003). Their approach is based on the semi-parametric method: the Cox proportional hazard model. Equation (4.3.22) can be regarded as a parametric survival model allowing both noncompliance and a non-ignorable censoring mechanism. This is comparable to the Frangakis-Rubin method (Frangakis & Rubin, 1999), a non-parametric method, which accommodates noncompliance and a non-ignorable censoring mechanism. It is noteworthy that the *latent ignorability* assumption is required in both of these two models, equation (4.3.22) and the Frangakis-Rubin model (Frangakis & Rubin, 1999).

Chapter 5

Parametric potential-outcome survival models

5.1 Motivation

Estimating the causal effect of interventions is often required when undertaking empirical studies in medicine. Randomisation is generally the only accepted approach for causal inference, but randomised controlled trials (RCTs) often suffer from noncompliance (Mealli et al., 2004), because not all participants follow their assigned treatment (Z_i) properly in practice. As stated in Section 1.7.2, compliance type U_i is often used to describe this situation. By its definition, compliance type U_i is determined by the potential received treatment $\mathbf{D}_i = [D_i(1), D_i(0)]$ (equation (1.7.2)). Since only one of $D_i(1)$ or $D_i(0)$ is observed in a RCT, compliance type U_i is an unobservable variable.

Many RCTs in medical field involve time-to-event outcome(s) of interest; often, it is the time to a predefined event. *Censoring* is also a primary feature of time-to-event data (Chapter 3). The type of censoring mechanism is important, in that it will determine the structure of the likelihood function (Chapter 4).

More specifically, in Section 4.2, we have argued that if compliance type U_i, an unobservable variable, is involved in a time-to-event study (such as in a RCT with noncompliance), in general, the censoring mechanism should be treated as non-ignorable. This is true unless the case of no loss to follow-up or the assumption of noninformative censoring is given as in Loeys & Goetghebeur (2003). When the censoring mechanism is not ignorable, extra assumptions are required. These are detailed in Section 4.3.3 of Chapter 4.

In Chapter 2, we reviewed current potential-outcome models that are pertinent to general data (Section 2.3.1) or to time-to-event data (Section 2.3.2). It was shown that among the current potential-outcome models for time-to-event data (Section 2.3.2), few consider noncompliance and non-response (censoring mechanism) simultaneously. It is noteworthy that except for Frangakis & Rubin (1999), few models for time-to-event data have been derived. On the other hand, for

the few models that do consider noncompliance and non-response simultaneously, whilst they follow the assumption of *latent ignorability*, they focus only on general data (Yau & Little, 2001; Mealli et al., 2004; Peng et al., 2004; O'Malley & Normand, 2005), and do not consider time-to-event data.

In this chapter, we follow the assumptions (Section 5.2) as given in Frangakis & Rubin (1999) and aim to establish a parametric potential-outcome (PPO) survival model for RCTs with *all-or-none compliance* for the case where censoring is allowed, namely a non-ignorable censoring mechanism parametric potential-outcome (PPO) survival model, or the non-ignorable PPO-survival model for short (Section 5.3 - Section 5.9). Furthermore, the ignorable censoring mechanism parametric potential-outcome (PPO) survival model or an ignorable PPO-survival model for short, is addressed in Section 5.10. For the ignorable PPO-survival model, an extra assumption is required (Section 5.10.1). To the best of the author's knowledge, there has not been any work to date that addresses both treatment noncompliance and non-response (missing) simultaneously in a time-to-event scenario.

The remains of this chapter follows this outline: Section 5.2 defines all assumptions required for creating the PPO-survival models. The compliance type variable and its distribution is then discussed in Section 5.3. Section 5.4 structures the joint distribution of all potential observable variables. The likelihood function of the PPO-survival model is given in Section 5.5, and they are specified to Weibull and log-normal distributions for failure and censoring time in Section 5.6. Section 5.7 addresses the AFT model and its property in Weibull or log-normal distribution. Section 5.8 appoints the interest parameters for specific PPO-survival models. The issues surrounding the model computation are given in Section 5.9. The special case of the PPO-survival model, the so-called ignorable PPO-survival model, is disccused in Section 5.10. The summary of this chapter is given in Section 5.11.

5.2 Assumptions

Recall the assumptions given in Frangakis & Rubin (1999). They are as follows:

1. The stable unit treatment value assumption (SUTVA) (Rubin, 1986; Imbens & Rubin, 1997; Peng et al., 2004): this assumes that the potential outcomes of each individual are unrelated to the treatment status of other individuals.

2. Random assignment: this assumes that each individual has been assigned randomly to a treatment group.

3. Exclusion restriction: this assumes that the assigned treatment has an effect on the outcome only through its effect on the received treatment.

4. Monotonicity: the actual received treatment $D_i(z)$ is a monotone function with respect to the potential treatment levels, 1 for being in the treatment group and 0 for being in the control group.

5. Latent ignorability: this assumes that the potential outcomes (of failure time) and the associated potential non-responses (censoring time) are independent at each level of the latent compliance types.

In the literature, the *latent ignorability* assumption for general data from RCTs has been expressed as follows:

$$Pr(\mathbf{Y}_i, \mathbf{R}_i|U_i, Z_i) = Pr(\mathbf{Y}_i|U_i, Z_i)Pr(\mathbf{R}_i|U_i, Z_i); \tag{5.2.1}$$

where $\mathbf{Y}_i$ and $\mathbf{R}_i$ represent the potential-outcome variable and the response indicator for the *ith* individual, respectively. The notation we used here, U_i and Z_i are the same as in Frangakis & Rubin (1999), Mealli et al. (2004) and O'Malley & Normand (2005). That is, U_i denotes the compliance type, and Z_i indicates the assigned treatment.

The analogous assumption for time-to-event data is:

$$Pr(\mathbf{T}_i, \mathbf{C}_i|U_i, Z_i) = Pr(\mathbf{T}_i|U_i, Z_i)Pr(\mathbf{C}_i|U_i, Z_i), \tag{5.2.2}$$

where $\mathbf{T}_i$ and $\mathbf{C}_i$ denote the potential outcomes vector for failure time and censoring time respectively. This assumes that in RCTs with time-to-event outcomes, the potential failure and censoring times are independent conditional on the latent compliance type. We regard equation (5.2.2) as the time-to-event version of the *latent ignorabilty* assumption given by Frangakis & Rubin (1999). Note that in this chapter, we use the bold symbols, such as $\mathbf{T}_i$ and $\mathbf{C}_i$, to denote potential-outcome vectors.

5.3 Compliance type and its distribution

Noncompliance is the main issue we have to consider in our PPO-survival approach and it is often described by compliance type variable, which is defined by the vector of potential received treatment $\mathbf{D}_i = [D_i(1),\ D_i(0)]$. As discussed in Section 1.7.2, *compliers* are defined by $D_i(1) = 1$ and $D_i(0) = 0$; *alawys-takers* are defined

by $D_i(1) = 1$ and $D_i(0) = 1$; and *never-takers* are defined by $D_i(1) = 0$ and $D_i(0) = 0$. By following the property of conditional probability, $Pr(\mathbf{D}_i|Z_i, \varepsilon)$, the conditional distribution of vector $\mathbf{D}_i$ given the assigned treatment Z_i, (where ε is the associated parameter for this distribution) is broken down in equation (5.3.1). This is because $\mathbf{D}_i$ can also be expressed as $[D_i^{mis},\ D_i^{obs}]$. Recall that in Section 1.7.1, we have defined $D_i^{mis} = D_i(1 - Z_i)$ and $D_i^{obs} = D_i(Z_i)$, which denote the unobservable and observable components of potential received treatment vector $\mathbf{D}_i$.

$$\begin{aligned} Pr(\mathbf{D}_i|Z_i;\varepsilon) &= Pr(D_i^{mis},\ D_i^{obs}|Z_i;\varepsilon) \\ &= Pr(D_i^{mis}|D_i^{obs}, Z_i;\varepsilon_{zd})Pr(D_i^{obs}|Z_i;\varepsilon_z), \end{aligned} \tag{5.3.1}$$

where ε_{zd} and ε_z represent the parameters associated with these two conditional distributions of D_i^{mis} and D_i^{obs} respectively.

On the other hand, by the definition of compliance type U_i, given the actual received treatment $D_i^{obs} = D_i(Z_i)$, learning the value of D_i^{mis} is equivalent to learning the value of U_i. This implies that the conditional probability distribution of the compliance type U_i given $D_i^{obs} = D_i(Z_i)$ and Z_i, $Pr(U_i|D_i(Z_i), Z_i, ;\boldsymbol{\pi})$, can be obtained from $Pr(D_i^{mis}|D_i^{obs}, Z_i;\varepsilon)$, the conditional probability distribution of D_i^{mis}, the unobservable component of the potential received treatment vector $\mathbf{D}_i$, given the actual received treatment D_i^{obs} and assigned treatment Z_i. In other words:

$$Pr(U_i|D_i^{obs}, Z_i;\boldsymbol{\pi}) \Leftarrow^{d}\Rightarrow Pr(D_i^{mis}|D_i^{obs}, Z_i;\varepsilon_{zd}); \tag{5.3.2}$$

where $\Leftarrow^{d}\Rightarrow$ indicates distributional equivalence: any probability on one side of the equation can be obtained from the distribution on the other side. For example, $Pr(U_i = c|D_i^{obs} = 1, Z_i = 1, \boldsymbol{\pi}) = Pr(D_i^{mis} = 0|D_i^{obs} = 1, Z_i = 1, \varepsilon_{zd})$.

After substituting equation (5.3.2) into the second line in equation (5.3.1), we have:

$$Pr(\mathbf{D}_i|Z_i, \varepsilon) \Leftarrow^{d}\Rightarrow Pr(U_i|D_i^{obs}, Z_i;\boldsymbol{\pi})Pr(D_i^{obs}|Z_i;\varepsilon_z), \tag{5.3.3}$$

where $\boldsymbol{\pi}$ denotes the parameter corresponding to the distribution of U_i conditional on the actual received treatment D_i^{obs} and assigned treatment Z_i, and ε_z is the parameter associated with the distribution of D_i^{obs} given Z_i. This implies that $Pr(\mathbf{D}_i|Z_i, \varepsilon)$ can be obtained from the product of $Pr(U_i|D_i^{obs}, Z_i;\boldsymbol{\pi})$ and $Pr(D_i^{obs}|Z_i;\varepsilon_z)$. In other words, $Pr(\mathbf{D}_i|Z_i, \varepsilon)$, the conditional distribution of $\mathbf{D}_i$ (the potential received treatment vector) given assigned treatment Z_i, can also be obtained from the product of two conditional distributions, $Pr(U_i|D_i^{obs}, Z_i;\boldsymbol{\pi})$ and $Pr(D_i^{obs}|Z_i;\varepsilon_z)$, the former is the conditional distribution of U_i given the actual re-

ceived treatment D_i^{obs} and assigned treatment Z_i, and the latter is the conditional distribution of D_i^{obs} given Z_i (equation (5.3.3)).

5.4 The structure of the joint distribution

5.4.1 The joint distribution of $\mathbf{T}$, $\mathbf{C}$, $\mathbf{D}$ *and* $\mathbf{Z}$

Suppose a RCT includes N individuals, each component of the potential outcomes vector, including failure time $\mathbf{T}_i = [T_i(1), T_i(0)]$, censoring time $\mathbf{C}_i = [C_i(1), C_i(0)]$ and the potential received treatment $\mathbf{D}_i = [D_i(1), D_i(0)]$, can then be written as a vector of length N, as follows:

$\underline{\mathbf{T}}(1)$ $= [T_1(1), \ \ldots, \ T_N(1)]$, the failure time if assigned to the treatment group;

$\underline{\mathbf{T}}(0)$ $= [T_1(0), \ \ldots, \ T_N(0)]$, the failure time if assigned to the control group;

$\underline{\mathbf{C}}(1)$ $= [C_1(1), \ \ldots, \ C_N(1)]$, the censoring time if assigned to the treatment group;

$\underline{\mathbf{C}}(0)$ $= [C_1(0), \ \ldots, \ C_N(0)]$, the censoring time if assigned to the control group;

$\underline{\mathbf{D}}(1)$ $= [D_1(1), \ \ldots, \ D_N(1)]$, the received treatment if assigned to the treatment group;

$\underline{\mathbf{D}}(0)$ $= [D_1(0), \ \ldots, \ D_N(0)]$, the received treatment if assigned to the control group;

We also assume that all related variables are independent and identically distributed (i.i.d.) random variables, which include failure time $\underline{\mathbf{T}}$, censoring time $\underline{\mathbf{C}}$, potential received treatment $\underline{\mathbf{D}}$ and assigned treatment $\underline{\mathbf{Z}}$. Their joint distribution can then be expressed as a product of the joint distribution for N single observations, that is:

$$\begin{aligned} &Pr(\underline{\mathbf{T}}(1), \underline{\mathbf{T}}(0), \underline{\mathbf{C}}(1), \underline{\mathbf{C}}(0), \underline{\mathbf{D}}(1), \underline{\mathbf{D}}(0), \underline{\mathbf{Z}}; \boldsymbol{\Psi}) \\ &\quad = \prod_i Pr(T_i(1), T_i(0), C_i(1), C_i(0), D_i(1), D_i(0), Z_i; \boldsymbol{\Psi}), \end{aligned} \tag{5.4.1}$$

where $\boldsymbol{\Psi}$ denotes the model parameter vector.

We suppose that the model parameter $\boldsymbol{\Psi}$ can be broken down into three parts, namely $\boldsymbol{\Psi} = (\boldsymbol{\theta}, \varepsilon, \chi)$, where $\boldsymbol{\theta}$ is the parameter associated with the joint distribution of failure and censoring times (outcomes), ε is the parameter for the distribution of the vector of potential received treatment $\mathbf{D}_i$ (post-treatment covariates), and χ is the parameter corresponding to the distribution of assigned

treatment Z_i (pre-treatment covariates). According to the rule of conditional distribution, the joint probability of all the aforementioned variables, corresponding to the *ith* individual, $Pr(T_i(1), T_i(0), C_i(1), C_i(0), D_i(1), D_i(0), Z_i; \mathbf{\Psi})$, can thus be expressed as a product of three distributions:

- The joint distribution of failure time $\mathbf{T}_i = [T_i(1), T_i(0)]$ and censoring time $\mathbf{C}_i = [C_i(1), C_i(0)]$ given potential received treatment $\mathbf{D}_i = [D_i(1), D_i(0)]$ and assigned treatment Z_i, $Pr(T_i(1), T_i(0), C_i(1), C_i(0)|D_i(1), D_i(0), Z_i; \boldsymbol{\theta})$.
- The distribution of potential received treatment $\mathbf{D}_i = [D_i(1), D_i(0)]$ given assigned treatment Z_i, $Pr(D_i(1), D_i(0)|Z_i; \varepsilon)$.
- The distribution of assigned treatment Z_i, $Pr(Z_i, \chi)$.

Note that $\boldsymbol{\theta}$, ε and χ denote the parameters associated with these above three distributions, respectively. We then have:

$$\begin{aligned} &Pr(T_i(1), T_i(0), C_i(1), C_i(0), D_i(1), D_i(0), Z_i; \mathbf{\Psi}) = \\ &\quad Pr(T_i(1), T_i(0), C_i(1), C_i(0)|D_i(1), D_i(0), Z_i; \boldsymbol{\theta}) Pr(D_i(1), D_i(0)|Z_i; \varepsilon) Pr(Z_i; \chi). \end{aligned} \tag{5.4.2}$$

Using the notation of potential-outcome vectors, equation (5.4.2) can thus be expressed as follows:

$$\begin{aligned} &Pr(\mathbf{T}_i, \mathbf{C}_i, \mathbf{D}_i, Z_i; \mathbf{\Psi}) = \\ &\qquad Pr(\mathbf{T}_i, \mathbf{C}_i|\mathbf{D}_i, Z_i; \boldsymbol{\theta}) Pr(\mathbf{D}_i|Z_i; \varepsilon) Pr(Z_i; \chi); \end{aligned} \tag{5.4.3}$$

for $\mathbf{\Psi} = (\boldsymbol{\theta}, \varepsilon, \chi)$, the model parameter vector defined earlier.

In Section 5.3, we derived equation (5.3.2), which allows us to replace $Pr(\mathbf{D}_i|Z_i; \varepsilon)$ with $Pr(U_i|D_i^{obs}, Z_i; \pi) Pr(D_i^{obs}|Z_i; \varepsilon_z)$. The joint distribution of all the aforementioned variables given in equation (5.4.1) can now be rewritten as follows:

$$\begin{aligned} &Pr(\mathbf{T}_i, \mathbf{C}_i, \mathbf{D}_i, Z_i; \Psi) \Leftarrow^{d} \Rightarrow \\ &\qquad Pr(\mathbf{T}_i, \mathbf{C}_i|Z_i, \mathbf{D}_i; \boldsymbol{\theta}) Pr(U_i|Z_i, D_i^{obs}; \pi) Pr(D_i^{obs}|Z_i; \varepsilon_z) Pr(Z_i; \chi), \\ &\qquad = Pr(\mathbf{T}_i, \mathbf{C}_i|Z_i, \mathbf{D}_i; \boldsymbol{\theta}) Pr(U_i|Z_i, D_i^{obs}; \pi) Pr(D_i^{obs}, Z_i; \lambda); \end{aligned} \tag{5.4.4}$$

where $\Leftarrow^{d} \Rightarrow$ indicates equivalence in distribution.

The last equality holds in equation (5.4.4) is because

$$Pr(D_i^{obs}, Z_i; \lambda) = Pr(D_i^{obs}|Z_i, \varepsilon_z) Pr(Z_i; \chi). \tag{5.4.5}$$

Therefore, these two parameters ε_z and χ, have been replaced with the parameter of λ. The advantage of using λ to replace ε_z and χ is that it simplifies notation of the parameters involved in the model. Recall that the main parameter of interest in our approach is $\boldsymbol{\theta}$, the parameter associated with the joint distribution of failure time and censoring time (survival outcomes).

In addition, the vector of received treatment $\mathbf{D}_i$ can be viewed as being equivalent to compliance type U_i. This is because, on one hand, $U_i = u$ (for $u \in \{c, a, n, d\}$) is determined by $\mathbf{D}_i = [D_i(1) = d_1, D_i(0) = d_0]$ (for d_0, $d_1 \in \{0, 1\}$) (equation (1.7.2)), and, on the other hand, if we know the compliance type $U_i = u$, the values of vector $\mathbf{D}_i = [D_i(1), D_i(0)]$ are then also known. So we write $(U_i = u) \Leftrightarrow (\mathbf{D}_i = [D_i(1), D_i(0)] = [d_1, d_0])$ where $\Leftrightarrow$ denotes equivalence.

Let U_i replace $\mathbf{D}_i$. We have:

$$Pr(\mathbf{T}_i, \mathbf{C}_i, \mathbf{D}_i, Z_i; \Psi) \Leftarrow^{d} \Rightarrow Pr(\mathbf{T}_i, \mathbf{C}_i, U_i, Z_i; \Psi); \tag{5.4.6}$$

then equation (5.4.4) is equivalent to

$$\begin{aligned} &Pr(\mathbf{T}_i, \mathbf{C}_i, U_i, Z_i; \Psi) \\ &\Leftarrow^{d} \Rightarrow Pr(\mathbf{T}_i, \mathbf{C}_i | Z_i, U_i; \boldsymbol{\theta}) Pr(U_i | Z_i, D_i^{obs}; \pi) Pr(D_i^{obs}, Z_i; \lambda). \end{aligned} \tag{5.4.7}$$

Furthermore, it is useful to impose some additional structure on the model. This is achieved by partitioning the parameter vector as $\boldsymbol{\theta} = (\boldsymbol{\theta}_T, \boldsymbol{\theta}_C, \theta_{TC})^T$, where ^{T} denotes transpose. We also require that the marginal bivariate failure time distribution depends only on $\boldsymbol{\theta}_T$ and that the marginal bivariate censoring time distribution depends only on $\boldsymbol{\theta}_C$. Hence the θ_{TC} parameters act as association parameters. This latter condition is only required to complete the parametrisation of the joint distribution of the potential failure and censoring times.

By following the assumption of *latent ignorability*, equation (5.2.2), which gives the joint distribution of failure and censoring times conditional on U_i and Z_i, denoted by $Pr(\mathbf{T}_i, \mathbf{C}_i | Z_i, U_i; \boldsymbol{\theta})$, can be separated as a product of two distributions of the failure time and the censoring time (conditional on compliance type U_i only because of randomisation). Specifically:

$$Pr(\mathbf{T}_i, \mathbf{C}_i | U_i, Z_i; \boldsymbol{\theta}) = Pr(\mathbf{T}_i | U_i, Z_i; \boldsymbol{\theta}_T) Pr(\mathbf{C}_i | U_i, Z_i; \boldsymbol{\theta}_C), \tag{5.4.8}$$

This means we can break down $\boldsymbol{\theta}$ into $\boldsymbol{\theta} = (\boldsymbol{\theta}_T, \boldsymbol{\theta}_C)^T$. In other words, under the *latent ignorability* assumption, the third part of $\boldsymbol{\theta}$, namely θ_{TC}, has been omitted.

Consequently, the model parameter Ψ can then be broken down as $\Psi = \{\boldsymbol{\theta}_T, \boldsymbol{\theta}_C, \boldsymbol{\pi}, \lambda\}$, where the subcomponent parameters of Ψ are detailed below:

$\boldsymbol{\theta}_T$ corresponds to the parametrisation of the distribution of failure time,

$\boldsymbol{\theta}_C$ corresponds to the parametrisation of the distribution of censoring time,

$\boldsymbol{\pi}$ corresponds to the parametrisation of the conditional distribution of compliance type U_i, and

λ corresponds to the parametrisation of the joint distribution of (D_i^{obs}, Z_i).

Equation (5.4.8) can also be expressed as the following equivalence relationships:

$$\begin{aligned}\Pr(T_i(1), C_i(1)|U_i, Z_i = 1; \boldsymbol{\theta}) =& \Pr(T_i(1)|U_i, Z_i = 1; \theta_{T1})\Pr(C_i(1)|U_i, Z_i = 1; \theta_{C1}) \\ \Pr(T_i(0), C_i(0)|U_i, Z_i = 0; \boldsymbol{\theta}) =& \Pr(T_i(0)|U_i, Z_i = 0; \theta_{T0})\Pr(C_i(0)|U_i, Z_i = 0; \theta_{C0}).\end{aligned} \tag{5.4.9}$$

We propose a similar structuring of the marginal bivariate failure and censoring time models by assuming $\boldsymbol{\theta}_T = (\theta_{T1}, \theta_{T0}, \theta_{T10})$ and $\theta_C = (\theta_{C1}, \theta_{C0}, \theta_{C10})$ respectively. We then have:

$$\Pr(T_i(0)|U_i; \boldsymbol{\theta}_T) = \Pr(T_i(0)|U_i; \theta_{T0}) \tag{5.4.10}$$

$$\Pr(T_i(1)|U_i; \boldsymbol{\theta}_T) = \Pr(T_i(1)|U_i; \theta_{T1}). \tag{5.4.11}$$

$$\Pr(C_i(0)|U_i; \boldsymbol{\theta}_C) = \Pr(C_i(0)|U_i; \theta_{C0}) \tag{5.4.12}$$

$$\Pr(C_i(1)|U_i; \theta_C) = \Pr(C_i(1)|U_i; \theta_{C1}). \tag{5.4.13}$$

Note that the third components of $\boldsymbol{\theta}_T$ and $\boldsymbol{\theta}_C$, denoted by θ_{T10} and θ_{C10}, were not involved. This can be interpreted as that an individual in a RCT can only be assigned into one of treatment groups ($Z_i = 1$ or 0). No one can be assigned into both two treatment groups ($Z_i = 1$ and 0). In other words, in practice no such a situation exists which need parameter θ_{T10} or θ_{C10} to describe. So we simplify $\boldsymbol{\theta}_T$ and $\boldsymbol{\theta}_C$ as $\boldsymbol{\theta}_T = (\theta_{T1}, \theta_{T0})$ and $\boldsymbol{\theta}_C = (\theta_{C1}, \theta_{C0})$.

Under randomisation, the assigned treatment Z is independent of all other variables. This is denoted as $Z \perp (\mathbf{T}, \mathbf{C}, \mathbf{D}, U)$, where $\perp$ means independent. It therefore follows that:

$$Pr(\mathbf{T}_i, \mathbf{C}_i|Z_i, U_i; \boldsymbol{\theta}) = Pr(\mathbf{T}_i, \mathbf{C}_i|U_i; \boldsymbol{\theta}). \tag{5.4.14}$$

Equation (5.4.14) represents a major simplification because it means that the dependency of the potential survival (and censoring) times on assigned treatment (Z_i) does not have to be explicitly modelled. In non-randomised studies, this dependency normally has to be modelled. Without additional assumptions, this generally leads to problems with identifiability (Rubin, 1978; Graham, 2000).

The joint distribution of all potential entities for the *ith* observation can then be expressed as follows:

$$\begin{aligned} & Pr(\mathbf{T}_i,\mathbf{C}_i,\mathbf{D}_i,Z_i;\Psi) \\ \Leftarrow^{d}\Rightarrow\ & Pr(\mathbf{T}_i|U_i;\boldsymbol{\theta}_T)Pr(\mathbf{C}_i|,U_i;\boldsymbol{\theta}_C)Pr(U_i|D_i^{obs};\boldsymbol{\pi})pr(D_i^{obs},Z_i;\lambda); \end{aligned} \tag{5.4.15}$$

where the model parameter Ψ is $\Psi = \{\boldsymbol{\theta}_T, \boldsymbol{\theta}_C, \boldsymbol{\pi}, \lambda\}$.

Our primary focus is on $\boldsymbol{\theta}_T$. In particular, we wish to focus on $(\theta_{T1}, \theta_{T0})$, the parameters of the marginal failure time distributions. Given estimates of these parameters and given our specific distributional assumptions, the estimated potential survivor functions (under the alternative assigned treatment) and the estimators of ACE(t) and CACE(t) can then be obtained. Here ACE(t) and CACE(t) are the forms of *average causal effect* (ACE) and *complier average causal effect* (CACE) (Angrist et al., 1996; Imbens & Rubin, 1997; Yau & Little, 2001; Peng et al., 2004; Mattei & Mealli, 2007), respectively, for time-to-event studies. Note that they both are functions of time t.

5.4.2 The joint distribution of the observable variables

For deriving the formulation of the likelihood function, we need to address the construct of the joint distribution of observed variables first, as it is well known that the likelihood function of the model parameter Ψ, denoted by $L^F(\Psi)$, is a function of Ψ, for fixed observed variables, and the function is proportional to the joint distribution of all observed variables.

In time-to-event studies, however, failure time T_i and censoring time C_i cannot be fully observed jointly. It is convenient to present a construct of the joint distribution of all observable variables (denoted by $Pr(\underline{T}^{obs}, \underline{C}^{obs}, \underline{D}^{obs}, \underline{Z}; \Psi)$), rather than the joint distribution of all observed variables, as the former can be obtained directly by integrating $Pr(\mathbf{T}_i, \mathbf{C}_i, \mathbf{D}_i, Z_i; \Psi))$ with respect to the unobservable/missing variables, which are $T_i^{mis} = T_i(1 - Z_i)$, $C_i^{mis} = C_i(1 - Z_i)$ and $D_i^{mis} = D_i(1 - Z_i)$ when Z_i is a binary. Therefore:

$$
\begin{aligned}
&Pr(T_i^{obs}, C_i^{obs}, D_i^{obs}, Z_i; \Psi) \\
&\quad = \int\int\int Pr(\mathbf{T}_i, \mathbf{C}_i, \mathbf{D}_i, Z_i; \Psi) dD_i^{mis} dC_i^{mis} dT_i^{mis},
\end{aligned}
\tag{5.4.16}
$$

where $Pr(\mathbf{T}_i, \mathbf{C}_i, \mathbf{D}_i, Z_i; \Psi)$ is given in equation (5.4.15).

Because of equation (5.4.15), equation (5.4.16) can then be expressed as follows

$$
\begin{aligned}
&Pr(T_i^{obs}, C_i^{obs}, D_i^{obs}, Z_i; \Psi) \Leftarrow^{d} \Rightarrow \\
&\int\int\int Pr(\mathbf{T}_i|U_i; \boldsymbol{\theta}_T) Pr(\mathbf{C}_i|U_i; \boldsymbol{\theta}_C) Pr(U_i|D_i^{obs}; \boldsymbol{\pi}) Pr(D_i^{obs}, Z_i; \lambda) dU_i dC_i^{mis} dT_i^{mis},
\end{aligned}
\tag{5.4.17}
$$

where dD_i^{mis} is replaced by dU_i because, given D_i^{obs} and Z_i, we have $D_i^{mis} \Leftrightarrow U_i$. Recall that $D_i^{obs} = D_i(Z_i)$, this implies that for the *ith* individual, his/her assigned treatment Z_i and actual received treatment D_i^{obs} are always observed.

Table 5.1: Compliance types $U_i = u$ given observed values of Z_i and $D_i(Z_i)$.

Z_i	$D_i(Z_i)$	Possible U_i
1	1	a or c
1	0	n
0	1	a
0	0	n or c

Under the assumptions of *exclusion restriction* and *monotonicity*, compliance type U_i has up to three categories, which are compliers ($U_i = c$), never-takers ($U_i = n$) and always-takers ($U_i = a$) and shown in Table 5.1. Then integration in equation (5.4.17), with respect to U_i (compliance type), can be further replaced by its summation. We then write:

$$
\begin{aligned}
&Pr(T_i^{obs}, C_i^{obs}, D_i^{obs}, Z_i; \Psi) \Leftarrow^{d} \Rightarrow \\
&\int\int \sum_{U_i \in \{c,n,a\}} Pr(\mathbf{T}_i|U_i; \boldsymbol{\theta}_T) Pr(\mathbf{C}_i|U_i; \boldsymbol{\theta}_C) Pr(U_i|D_i^{obs}; \boldsymbol{\pi}) Pr(D_i^{obs}, Z_i; \lambda) dC_i^{mis} dT_i^{mis}, \\
&= Pr(D_i^{obs}, Z_i; \lambda) \sum_{U_i \in \{c,n,a\}} Pr(U_i|D_i^{obs}; \boldsymbol{\pi}) \int Pr(\mathbf{T}_i|U_i; \boldsymbol{\theta}_T) dT_i^{mis} \int Pr(\mathbf{C}_i|U_i; \boldsymbol{\theta}_C) dC_i^{mis}, \\
&= Pr(D_i^{obs}, Z_i; \lambda) \sum_{U_i \in \{c,n,a\}} Pr(T_i^{obs}|U_i; \boldsymbol{\theta}_T) Pr(C_i^{obs}|U_i; \boldsymbol{\theta}_C) Pr(U_i|D_i^{obs}; \boldsymbol{\pi}).
\end{aligned}
\tag{5.4.18}
$$

where $\Leftarrow\!\!^{d}\!\Rightarrow$ denotes distributional equivalence.

5.5 The likelihood function

As stated in Section 5.4.2, the likelihood function with respect to the model parameter $\mathbf{\Psi}$ can be a function that is proportional to the joint distribution of all observed variables (p9, Tanner (1993)). However, because in time-to-event studies for the *ith* individual, either failure time (T_i^{obs}) or censoring time (C_i^{obs}) can be observed, but not both. This implies that $Pr(T_i^{obs}, C_i^{obs}, D_i^{obs}, Z_i; \mathbf{\Psi})$ can only be viewed as the joint distribution for the observable variables (equation (5.4.18)) rather than the joint distribution for the observed variables.

Hence, instead of the likelihood function, we formulate a function of the model parameter Ψ that is proportional to the joint distribution of all observable variables, the so-called the idealised likelihood function with respect to the model parameter $\mathbf{\Psi}$ denoted by $IDL(\mathbf{\Psi})$.

5.5.1 The idealised likelihood function

Given the joint distribution of the observable variables, $Pr(T_i^{obs}, C_i^{obs}, D_i^{obs}, Z_i|\Psi)$, in equation (5.4.18), for the *ith* observation, the idealised likelihood function with respect to the model parameter Ψ, denoted by $IDL_i(\Psi)$, has the following form:

$$IDL_i(\Psi) \propto Pr(T_i^{obs}, C_i^{obs}, D_i^{obs}, Z_i; \Psi). \tag{5.5.1}$$

On the other hand, by following equation (5.4.18), the product of four distributions, $Pr(T_i^{obs}|U_i, Z_i; \boldsymbol{\theta}_T)$,$Pr(C_i^{obs}|U_i, Z_i; \boldsymbol{\theta}_C)$, $Pr(U_i|D_i^{obs}, Z_i; \pi)$ and $Pr(D_i^{obs}, Z_i; \lambda)$ can be viewed as a function of the model parameter $\Psi = (\boldsymbol{\theta}_T, \boldsymbol{\theta}_C, \boldsymbol{\pi}, \lambda)$, and it is proportional to the joint distribution of the idealised observed variables, $Pr(T_i^{obs}, C_i^{obs}, D_i^{obs}, Z_i; \Psi)$.

The idealised likelihood function with respect to the model parameter Ψ, ($IDL_i(\Psi)$), can then be obtained from the right hand side of equation (5.4.18), so

it has the formulation as follows:

$$
\begin{aligned}
&IDL_i(\Psi) \\
&\quad = \sum_{U_i \in \{c,n,a\}} Pr(T_i^{obs}|U_i, Z_i; \boldsymbol{\theta}_T) Pr(C_i^{obs}|U_i, Z_i; \boldsymbol{\theta}_C) Pr(U_i|D_i^{obs}, Z_i; \boldsymbol{\pi}) Pr(D_i^{obs}, Z_i; \lambda) \\
&\quad = Pr(D_i^{obs}, Z_i; \lambda) \sum_{U_i \in \{c,n,a\}} Pr(T_i^{obs}|U_i, Z_i; \boldsymbol{\theta}_T) Pr(C_i^{obs}|U_i, Z_i; \boldsymbol{\theta}_C) Pr(U_i|D_i^{obs}, Z_i; \boldsymbol{\pi}),
\end{aligned}
\tag{5.5.2}
$$

where the second equation holds because the term $Pr(D_i^{obs}, Z_i; \lambda)$ is not related to the compliance covariate U_i involved in the summation (equation (5.5.2)).

Note that in equation (5.5.2), assigned treatment Z_i is involved in each term of conditional distribution on the right-hand side. This is true because the observable variables, namely T_i^{obs}, C_i^{obs} and D_i^{obs}, all rely on the given value of Z_i.

Recall that in our approach, the main parameter of interest is $\boldsymbol{\theta}_T$, the parameter associated with the distribution of failure time T_i. However, from equation (5.5.2), we can see that because the three parameters, $\boldsymbol{\theta}_T$, $\boldsymbol{\theta}_C$ and $\boldsymbol{\pi}$, are all involved in the summation over compliance type U_i, the likelihood for these parameters does not separate into distinct components which can be maximised separately. This implies that estimation of $\boldsymbol{\theta}_T$ cannot proceed independently. So we write these three parameters as a row vector, $\boldsymbol{\beta} = (\boldsymbol{\theta}_T, \boldsymbol{\theta}_C, \boldsymbol{\pi})$, which are associated with the distributions of failure time, censoring time and compliance type, respectively. We then focus on deriving the idealised likelihood function for $\boldsymbol{\beta}$, denoted by $IDL_i(\boldsymbol{\beta})$ rather than $IDL_i(\boldsymbol{\Psi})$. Then $IDL_i(\boldsymbol{\beta})$ can be expressed as follows:

$$
\begin{aligned}
&IDL_i(\boldsymbol{\beta}) \\
&\quad = \sum_{U_i \in \{c,n,a\}} Pr(T_i^{obs}|U_i, Z_i; \boldsymbol{\theta}_T) Pr(C_i^{obs}|U_i, Z_i; \boldsymbol{\theta}_C) Pr(U_i|D_i^{obs}, Z_i; \boldsymbol{\pi}).
\end{aligned}
\tag{5.5.3}
$$

5.5.2 Likelihood function for complete data

In Section 4.2, the censoring indicator R_i was introduced to indicate the situation of observed failure time or censoring time, (1 for the event occurring

(observed failure time), 0 for censoring (observed censoring time)). Therefore, once given the censoring indicator R_i for the *ith* individual, the idealised likelihood function $IDL_i(\boldsymbol{\beta})$ can be specified to the likelihood function $L_i(\boldsymbol{\beta})$, which can be written as the following equation:

$$
\begin{aligned}
& L_i(\boldsymbol{\beta}) \\
= & \begin{cases}
\sum_{U_i \in \{c,n,a\}} \left\{ \begin{array}{l} pr(T_i^{obs} = t | U_i, Z_i; \boldsymbol{\theta}_T) \\ \times Pr(C_i^{obs} > t | U_i, Z_i; \boldsymbol{\theta}_C) \\ \times Pr(U_i | D_i^{obs}, Z_i; \pi) \end{array} \right\} & \text{for } R_i = 1,\ T_i(z) = t, \\
\sum_{U_i \in \{c,n,a\}} \left\{ \begin{array}{l} Pr(T_i^{obs} > t | U_i, Z_i; \boldsymbol{\theta}_T) \\ \times pr(C_i^{obs} = t | U_i, Z_i; \boldsymbol{\theta}_C) \\ \times Pr(U_i | D_i^{obs}, Z_i; \boldsymbol{\pi}) \end{array} \right\} & \text{for } R_i = 0,\ C_i(z) = t.
\end{cases}
\end{aligned}
\tag{5.5.4}
$$

where $Pr()$ and $pr()$ denote the probability function and the probability density function, respectively. However, in practice, the compliance type U_i cannot be observed fully. Hence the parameter $\boldsymbol{\beta}$ cannot be estimated through the form of $L_i(\boldsymbol{\beta})$ given in equation (5.5.4). Dempster et al. (1977) proposed an approach for MLE of incomplete data, named the EM-algorithm, which can be employed in our case as U_i is considered unknown and can be treated as incomplete data.

Briefly, the EM algorithm involves two steps: the E-step and the M-step, which denote the expectation and maximisation steps, respectively (Dempster et al., 1977). The E-step aims to obtain the conditional expectation of the log-likelihood function for complete data $\log(L^F(\boldsymbol{\beta}))$, denoted by $Q(\boldsymbol{\beta}, \boldsymbol{\beta}^k)$, given the observed data $\underline{T}(z)$, $\underline{C}(z)$, $\underline{D}(z)$ and $\underline{Z} = \underline{z}$, and given the current parameter values $\boldsymbol{\beta}^k$. More details for the EM algorithm is given in Section 5.5.5.

In the EM-algorithm, the first step requires the explication of the likelihood form for complete data denoted by $L^F(\boldsymbol{\beta})$. This depends on the assumption that all compliance covariates are known. We then find the related likelihood form for the *ith* observation, which can be written as follows (for

known compliance covariates $U_i = u$):

$$
\begin{aligned}
&L^F{}_i(\boldsymbol{\beta}) = L_i(\boldsymbol{\beta}, u) \\
&\quad = \left\{ \begin{array}{ll}
\left\{ \begin{array}{l} pr(T_i^{obs} = t|U_i, Z_i; \theta_T) \\ \times Pr(C_i^{obs} > t|U_i, Z_i; \theta_C) \\ \times Pr(U_i|D_i^{obs}, Z_i; \pi) \end{array} \right\} & \text{for } R_i = 1,\ T_i(z) = t, \\
\left\{ \begin{array}{l} Pr(T_i^{obs} > t|U_i, Z_i; \theta_T) \\ \times pr(C_i^{obs} = t|U_i, Z_i; \theta_C) \\ \times Pr(U_i|D_i^{obs}, Z_i; \pi) \end{array} \right\} & \text{for } R_i = 0,\ C_i(z) = t.
\end{array} \right.
\end{aligned} \tag{5.5.5}
$$

Compliance type $U_i = u$ is used to describe an individual's compliance behaviour. Therefore, we have redefined the parameter of interest, $\boldsymbol{\theta}_{T1} = (\theta_{T,u1})$ and $\boldsymbol{\theta}_{T0} = (\theta_{T,u0})$ (for $u \in \{c,\ a,\ n\}$). Recall that $\boldsymbol{\theta_T} = (\boldsymbol{\theta}_{T1},\ \boldsymbol{\theta}_{T0})$. More specifically, the parameter of interest associated with the failure time distribution, $\boldsymbol{\theta_T}$, is redefined as $\boldsymbol{\theta_T} = (\theta_{T,c1},\ \theta_{T,n1},\ \theta_{T,a1},\ \theta_{T,c0},\ \theta_{T,n0},\ \theta_{T,a0})$. This is equivalent to $\boldsymbol{\theta_T} = (\theta_{T,uz})$ (for $u \in \{a, c, n\}$ and $z \in \{0, 1\}$). Here the subscripts indicate failure time (T) with $U = u$ for $u \in \{a, c, n\}$ and assigned treatment $Z = z$ for $z \in \{0, 1\}$.

Similarly, we use $\theta_{C,uz}$ for $u \in \{a, c, n\}$, and $z \in \{0, 1\}$ to denote the components of $\boldsymbol{\theta}_C$, the parameter associated with the censoring time distribution. In other words, we have $\boldsymbol{\theta_C} = (\theta_{C,c1},\ \theta_{C,n1},\ \theta_{C,a1},\ \theta_{C,c0},\ \theta_{C,n0},\ \theta_{C,a0})$.

Hence, for the specific assigned treatment $Z_i = z$ (with known $U_i = u$), equation (5.5.5) can be simplified as follows:

$$
\begin{aligned}
&L_i(\boldsymbol{\beta}, u) \\
&\quad = \left\{ \begin{array}{ll}
\left\{ \begin{array}{l} pr(T(z)_i = t|U_i, Z_i; \theta_{T,uz}) \\ \times Pr(C(z)_i > t|U_i, Z_i; \theta_{C,uz}) \\ \times Pr(U_i|D_i(z), Z_i; \pi_{u,zd}) \end{array} \right\} & \text{for } R_i = 1,\ T_i(z) = t, \\
\left\{ \begin{array}{l} Pr(T(z)_i > t|U_i, Z_i; \theta_{T,uz}) \\ \times pr(C(z)_i = t|U_i, Z_i; \theta_{C,uz}) \\ \times Pr(U_i|D_i(z), Z_i; \pi_{u,zd}) \end{array} \right\} & \text{for } R_i = 0,\ C_i(z) = t.
\end{array} \right.
\end{aligned} \tag{5.5.6}
$$

where R_i is the observed indicator of censoring, which is 1 for the event of interest and 0 for censoring. The parameter $\pi_{u,zd}$ is associated with the

conditional probability of compliance, given $Z_i = z$ and $D_i(z) = d$, and is defined by $\pi_{u,zd} = Pr(U_i = u|Z_i = z, D_i(z) = d)$ for $u = (a, c, n)$ and $z,\ d = 0,\ 1$. Details about the estimation of the parameter $\pi_{u,zd}$ will be discussed in Section 5.9.

According to equation (5.5.6), an individual randomised to treatment $Z_i = 1$ who dies or is censored at (observed) time t has a contribution to the likelihood $\boldsymbol{\beta}$ as follows:

$$
\begin{aligned}
&L_{i\in\{i:Z_i=1\}}(\boldsymbol{\beta}, u) \\
&= \begin{cases}
\left\{ \begin{array}{l} pr(T(1)_i = t|U_i = u, Z_i = 1; \theta_{T,u1}) \\ \times Pr(C(1)_i > t|U_i = u, Z_i = 1; \theta_{C,u1}) \\ \times Pr(U_i|D_i(1), Z_i = 1; \pi_{u,1d}) \end{array} \right\} & \text{for } R_i = 1,\ T(1)_i = t, \\
\left\{ \begin{array}{l} Pr(T(1)_i > t|U_i = u, Z_i = 1; \theta_{T,u1}) \\ \times pr(C(1)_i = t|U_i = u, Z_i = 1; \theta_{C,u1}) \\ \times Pr(U_i|D_i(1), Z_i = 1; \pi_{u,1d}) \end{array} \right\} & \text{for } R_i = 0,\ C(1)_i = t.
\end{cases}
\end{aligned}
\tag{5.5.7}
$$

for $d = 0$ or 1 and $u \in \{c, a, n\}$.

Similarly, an individual randomised to treatment $Z_i = 0$ who dies or is censored at (observed) time t has a contribution to the likelihood for $\boldsymbol{\beta}$ as follows:

$$
\begin{aligned}
&L_{i\in\{i:Z_i=0\}}(\boldsymbol{\beta}, u) \\
&= \begin{cases}
\left\{ \begin{array}{l} pr(T(0)_i = t|U_i = u, Z_i = 0; \theta_{T,u0}) \\ \times Pr(C(0)_i > t|U_i = u, Z_i = 0; \theta_{C,u0}) \\ \times Pr(U_i|D_i(0), Z_i = 0; \pi_{u,0d}) \end{array} \right\} & \text{for } R_i = 1,\ T(0)_i = t, \\
\left\{ \begin{array}{l} Pr(T(0)_i > t|U_i = u, Z_i = 0; \theta_{T,u0}) \\ \times pr(C(0)_i = t|U_i = u, Z_i = 0; \theta_{C,u1}) \\ \times Pr(U_i|D_i(0), Z_i = 0; \pi_{u,0d}) \end{array} \right\} & \text{for } R_i = 0,\ C(0)_i = t.
\end{cases}
\end{aligned}
\tag{5.5.8}
$$

for $d = 0$ or 1 and $u \in \{c, n, a\}$.

Recall from Table 5.1 that knowing the values of $Z_i = z$ and $D_i(z) = d$ means that compliance type $U_i = u$ can be determined for up to two types. Hence, given the value of z and d, equations (5.5.7) and (5.5.8) can then be

expressed as the following form:

$$
\begin{aligned}
&L_{i\in\{i:Z_i=z,D_i(z)=d\}}(\boldsymbol{\beta};u) \\
&= \begin{cases}
\left\{\begin{array}{l} pr(T(z)_i=t|U_i=u,Z_i=z;\theta_{T,uz}) \\ \times Pr(C(z)_i>t|U_i=u,Z_i=z;\theta_{C,uz}) \\ \times Pr(U_i|D_i(z),Z_i=z;\pi_{u,zd}) \end{array}\right\} & \text{for } R_i=1,\ T(z)_i=t, \\
\left\{\begin{array}{l} Pr(T(z)_i>t|U_i=u,Z_i=z;\theta_{T,uz}) \\ \times pr(C(z)_i=t|U_i=u,Z_i=z;\theta_{C,uz}) \\ \times Pr(U_i|D_i(z),Z_i=z;\pi_{u,zd}) \end{array}\right\} & \text{for } R_i=0,\ C(z)_i=t.
\end{cases}
\end{aligned}
\tag{5.5.9}
$$

for z and $d = 0$ or 1 and $u \in \{c, a, n\}$. Note that equation (5.5.9) given above is a likelihood function of our non-ignorable PPO-survival model which is analogous to the one given in O'Malley & Normand (2005) model that is for normally distributed outcomes.

For notational convenience, we introduce the following conventions:

$$
\begin{aligned}
f^i_u(v,t;\theta_{T,uv}) &= pr(T_i(v)=t|U_i=u,\theta_{T,uv}) \quad \text{for } v=0 \text{ or } 1, \\
S^i_u(v,t;\theta_{T,uv}) &= Pr(T_i(v)>t|U_i=u,\theta_{T,uv}) \quad \text{for } v=0 \text{ or } 1, \\
e^i_u(v,t;\theta_{C,uv}) &= pr(C_i(v)=t|U_i=u,\theta_{C,uv}) \quad \text{for } v=0 \text{ or } 1, \\
G^i_u(v,t;\theta_{C,uv}) &= Pr(C_i(v)>t|U_i=u,\theta_{C,uv}) \quad \text{for } v=0 \text{ or } 1.
\end{aligned}
\tag{5.5.10}
$$

As mentioned before, $Pr()$ and $pr()$ denote the probability and probability density function, respectively and v indicates potential treatment levels.

Equation (5.5.9) can then be rewritten as:

$$
\begin{aligned}
&L_{i\in\{i:Z_i=z,D_i(z)=d\}}(\boldsymbol{\beta};u) \\
&= \left[f^i_u(z,t;\theta_{T,uz})G^i_u(z,t;\theta_{C,uz})\right]^{R_i}\left[S^i_u(z,t;\theta_{T,uz})e^i_u(z,t;\theta_{C,uz})\right]^{1-R_i}\pi_{u,zd}.
\end{aligned}
\tag{5.5.11}
$$

Furthermore, we write:

$$
\begin{aligned}
&F^i_{u,zd}(\theta_{T,uz},\theta_{C,uz}) \\
&\quad = \left[f^i_u(z,t;\theta_{T,uz})G^i_u(z,t;\theta_{C,uz})\right]^{R_i}\left[S^i_u(z,t;\theta_{T,uz})e^i_u(z,t;\theta_{C,uz})\right]^{1-R_i}.
\end{aligned}
\tag{5.5.12}
$$

Equation (5.5.9) can then be expressed more simply as:

$$L_{i\in\{i:Z_i=z,D_i(z)=d\}}(\boldsymbol{\beta};u)=F^i_{u,zd}(\theta_{T,uz},\theta_{C,uz})\pi_{u,zd}=l^i_\theta(\boldsymbol{\theta};u)l^i_\pi(\boldsymbol{\pi};u) \quad (5.5.13)$$

Because of the independent and identically distributed (i.i.d.) assumption for all observable variables (conditional on model parameters), the likelihood function for complete data denoted by $L^F(\beta)=L(\beta,u)$ can be expressed by the product of the fully likelihood function associated with each individual. In other words, we have:

$$L^F(\boldsymbol{\beta})=\prod_{i=1}L_i(\boldsymbol{\beta},u)=\prod_{i=1}F^i_{u,zd}(\theta_{T,uz},\theta_{C,uz})\pi_{u,zd}=\prod_{i=1}l^i_\theta(\boldsymbol{\theta};u)l^i_\pi(\boldsymbol{\pi};u). \quad (5.5.14)$$

Given observed $Z_i=z$ and $D_i(z)=d$, the likelihood form of θ for complete data, in which compliance type U_i (for each individual) is presumed to be known, can then be expressed as:

$$\begin{aligned}L^F(\boldsymbol{\beta})&=L(\boldsymbol{\beta},u)=\prod_i L_i(\boldsymbol{\beta},u)\\&=\begin{cases}\prod_{i\in\{i:U_i=c,Z_i=1,D_i(1)=1\}}F^i_{c,11}(\theta_{T,c1},\theta_{C,c1})\pi_{c,11}\\ \prod_{i\in\{i:U_i=a,Z_i=1,D_i(1)=1\}}F^i_{a,11}(\theta_{T,a1},\theta_{C,a})\pi_{a,11}\\ \prod_{i\in\{i:U_i=c,Z_i=0,D(0)=0\}}F^i_{c,00}(\theta_{T,c0},\theta_{C,c0})\pi_{c,00}\\ \prod_{i\in\{i:U_i=n,Z_i=0,D_i(0)=0\}}F^i_{n,00}(\theta_{T,n0},\theta_{C,n0})\pi_{n,00}\\ \prod_{i\in\{i:U_i=n,Z_i=1,D(0)=0\}}F^i_{n,10}(\theta_{T,n0},\theta_{C,n0})\pi_{n,10}\\ \prod_{i\in\{i:U_i=a,Z_i=0,D(0)=1\}}F^i_{a,01}(\theta_{T,a1},\theta_{C,a1})\pi_{a,01}.\end{cases}\end{aligned} \quad (5.5.15)$$

5.5.3 The likelihood form for incomplete data

Since compliance type U_i is unobservable, U_i needs to be treated as a variable with missing data. Hence, data associated with the case where noncompliance occurs have to be viewed as incomplete data.

Let $\mathcal{A}(z,d)$ denote the subset of subjects with specified values of Z and D; more specifically, $\mathcal{A}(z,d)=\{i:Z_i=z,\text{ and }D_i=d\}$. The likelihood form for incomplete data for the non-ignorable PPO-survival models can then be

written as shown below:

$$L(\boldsymbol{\beta}) = \sum_{u\in\{c,n,a\}} L(\boldsymbol{\beta}, u) = \sum_{u\in\{c,n,a\}} \prod_{i} L_i(\boldsymbol{\beta}, u)$$

$$= \begin{cases} \prod_{i\in\mathcal{A}(1,1)} \left\{ \left(F^i_{c,11}(\theta_{T,c1}, \theta_{C,c1})\pi_{c,11}\right) + \left(F^i_{a,11}(\theta_{T,a1}, \theta_{C,a1})\pi_{a,11}\right) \right\} \\ \prod_{i\in\mathcal{A}(0,0)} \left\{ \left(F^i_{c,00}(\theta_{T,c0}, \theta_{C,c0})\pi_{c,00}\right) + \left(F^i_{n,00}(\theta_{T,n0}, \theta_{C,n0})\pi_{n,00}\right) \right\} \\ \prod_{i\in\mathcal{A}(1,0)} \left(F^i_{n,10}(\theta_{T,n0}, \theta_{C,n0})\pi_{n,10}\right) \prod_{i\in\mathcal{A}(0,1)} \left(F^i_{a,01}(\theta_{T,a1}, \theta_{C,a1})\pi_{a,01}\right). \end{cases} \tag{5.5.16}$$

The likelihood function of $\boldsymbol{\beta}$ for incomplete data, given in equation (5.5.16), does not appear to be readily broken down into products of functions involving only θ_T and θ_C, respectively. That is $L(\boldsymbol{\beta}) \neq L(\theta_T)L(\theta_C)$. That is the case where censoring is not ignorable (Baker, 1997).

5.5.4 Simplification of model parameter $\boldsymbol{\beta}$

In equation (5.5.15), we observe that the parameter $\boldsymbol{\beta}$ has the form $\boldsymbol{\beta} = (\boldsymbol{\theta}_{T,uz}, \boldsymbol{\theta}_{C,uz}, \boldsymbol{\pi}_{u,zd})$ for $u = (a, c, n)$ and z, $d = (0, 1)$. However, the number of model parameters can be further reduced by following the assumption of *compound exclusion restriction* (Angrist et al., 1996; Frangakis & Rubin, 1999).

The compound exclusion restriction assumes that all potential-outcome vectors can be affected by the received treatment $\mathbf{D}_i(Z)$, rather than by the assigned treatment $\mathbf{Z}_i$. This implies that the outcome of interest for an individual, as long as his/her received treatment has the same value, will have the same value when he/she is assigned into either treatment group (if possible in practice). For convenience to express the assumption of *exclusion restriction* mathematically, we use the notation $T_i(Z_i = z, D_i(z) = d)$ and $C_i(Z_i = z, D_i(z) = d)$ to denote the *ith* individual's failure and censoring time where his/her assigned treatment Z_i and actual received treatment $D_i(Z_i)$ are given by $Z_i = z$ and $D_i(z) = d$, respectively. Recall that in a two-arm RCT, we have z and $d \in \{1, 0\}$. Hence the *exclusion restriction* assumption

can be expressed as follows:

$$\begin{aligned} &T_i(Z_i = 1, D_i(1) = d) = T_i(Z_i = 0, D_i(0) = d) \quad \text{for failure time,} \\ &C_i(Z_i = 1, D_i(1) = d) = C_i(Z_i = 0, D_i(0) = d) \quad \text{for censoring time.} \end{aligned} \tag{5.5.17}$$

This also implies that individuals who are never-takers or always-takers would have the same failure and censoring times probability distributions, whether they are in the treated group or in the control group. In other words:

$$\begin{aligned} &Pr(T_i(Z_i, D_i(Z_i))|Z_i = 0, U_i = u) \\ &\qquad = Pr(T_i(Z_i, D_i(Z_i))|Z_i = 1, U_i = u) \quad \text{for } \ u = a \, or \, n; \\ &Pr(C_i(Z_i, D_i(Z_i))|Z_i = 0, U_i = u) \\ &\qquad = Pr(C_i(Z_i, D_i(Z_i))|Z_i = 1, U_i = u) \quad \text{for } u = a \, or \, n. \end{aligned} \tag{5.5.18}$$

For convenience, in the remainder of this thesis, we use $T_i(Z_i)$ and $C_i(Z_i)$ to denote the failure and censoring times corresponding to the assigned treatment for the *ith* individual, whose assigned treatment is $Z_i = z$ and whose actual received treatment is $D_i(z) = d$. Equation (5.5.18) can then be rewritten as:

$$\begin{aligned} &Pr(T_i(0)|Z_i = 0, U_i = u) \\ &\qquad = Pr(T_i(1)|Z_i = 1, U_i = u) \quad \text{for } \ u = a \, or \, n; \\ &Pr(C_i(0)|Z_i = 0, U_i = u) \\ &\qquad = Pr(C_i(1)|Z_i = 1, U_i = u) \quad \text{for } u = a \, or \, n. \end{aligned} \tag{5.5.19}$$

Consequently, the parameters associated with the outcome distributions (failure and censoring), which correspond to the subgroup with $U_i = n$ and $Z_i = 1$ or 0, and to subgroup with $U_i = a$ and $Z_i = 1$ or 0, are equivalent. Hence, we have the following relationships:

$$\begin{aligned} \theta_{T,n1} &= \theta_{T,n0} \qquad \theta_{T,a1} = \theta_{T,a0} \\ \theta_{C,n1} &= \theta_{C,n0} \qquad \theta_{C,a1} = \theta_{T,a0}. \end{aligned} \tag{5.5.20}$$

Therefore we can use $\theta_{T,n}$ to represent both $\theta_{T,n1}$ and $\theta_{T,n0}$, as they are identical under the assumption of *exclusion restriction*. Similarly, the two

parameters $\theta_{T,a1}$ and $\theta_{T,a0}$ are identical under the same assumption. So we can use $\theta_{T,a}$ to denote both $\theta_{T,a1}$ and $\theta_{T,a0}$. By following the same logic, we write $\theta_{C,n} = \theta_{C,n1} = \theta_{C,n0}$ and $\theta_{C,a} = \theta_{C,a1} = \theta_{C,a0}$.

Equation (5.5.20) implies that the parameters $\theta_{T,uz}$ and $\theta_{C,uz}$, which are associated with the failure and censoring times for the subgroup with $U_i = u$ and assigned treatment $Z_i = z$, include the following components:

$$\begin{aligned} \boldsymbol{\theta}_{T,uz} &= (\theta_{T,c1},\ \theta_{T,c0},\ \theta_{T,n},\ \theta_{T,a}) \\ \boldsymbol{\theta}_{C,uz} &= (\theta_{C,c1},\ \theta_{C,c0},\ \theta_{C,n},\ \theta_{C,a}). \end{aligned} \tag{5.5.21}$$

Note that no parameter for the subgroup with $U_i = d$ and $Z_i = 1$ or 0 exists, because under the assumption of Monotonicity, no defiers exist.

Thereby, because we have assumed *monotonicity*, up to two categories exist for compliance type $U_i = u$ (knowing the value of Z_i and $D_i(Z_i)$) (Table 5.1). The parameter $\pi_{u,zd} = Pr(U_i = u|Z_i = z, D_i(z) = d)$, which is defined as the conditional probability of compliance type U_i given the assigned treatment $Z_i = z$ and actual received treatment $D^{obs} = d$, can now satisfy the following equations:

$$\begin{aligned} &\pi_{c,11} + \pi_{a,11} = 1 \quad \text{for } u \in \{c, a\} \text{ when } Z_i = 1 \text{ and } D_i(1) = 1, \\ &\pi_{c,00} + \pi_{n,00} = 1 \quad \text{for } u \in \{c, n\} \text{ when } Z_i = 0 \text{ and } D_i(0) = 0, \\ &\pi_{n,10} = 1 \qquad\quad \text{for } u \in \{n\} \quad \text{when } Z_i = 1 \text{ and } D_i(1) = 0, \\ &\pi_{a,01} = 1, \qquad\quad \text{for } u \in \{a\} \quad \text{when } Z_i = 0 \text{ and } D_i(0) = 1. \end{aligned} \tag{5.5.22}$$

The other $\pi_{u,zd}$ for $u \in \{c, n, a\}$ and $z,\ d = 1$ or 0 have zero values.

Now the model parameters of interest $\boldsymbol{\beta}$ can then be specified as $\boldsymbol{\beta} = (\boldsymbol{\theta}, \boldsymbol{\pi})$, where

$$\begin{aligned} \boldsymbol{\theta} &= (\boldsymbol{\theta}_{T,c1}, \boldsymbol{\theta}_{T,c0}, \boldsymbol{\theta}_{T,n}, \boldsymbol{\theta}_{T,a}, \boldsymbol{\theta}_{C,c1}, \boldsymbol{\theta}_{C,c0}, \boldsymbol{\theta}_{C,n}, \boldsymbol{\theta}_{C,a}), \\ \boldsymbol{\pi} &= (\pi_{c,11}, \pi_{c,00}). \end{aligned} \tag{5.5.23}$$

In addition, using π_u to denote the probability of compliance type $U_i = u$, i.e. $\pi_u = Pr(U_i = u)$, and by following the theorem of total probability, we

have:

$$\begin{aligned} Pr(U_i = u) &= \sum_{d\in\{1,0\}} \sum_{z\in\{1,0\}} Pr(U_i = u, D_i(z) = d, Z_i = z) \\ &= \sum_{d\in\{1,0\}} \sum_{v\in\{1,0\}} Pr(U_i = u|D_i(z) = d, Z_i = z)Pr(D_i(z) = d, Z_i = z). \end{aligned} \tag{5.5.24}$$

Recall that $\pi_{u,zd} = Pr(U_i = u|D_i(z) = d, Z_i = z)$. Denote $\lambda_{zd} = Pr(D_i(z) = d, Z_i = z)$ to correspond to the probability of individuals with values ($Z_i = z$, $D_i(z) = d$)) in the population. Then equation (5.5.24) can be re-written as:

$$\pi_u = \sum_{d\in\{1,0\}} \sum_{z\in\{1,0\}} \pi_{u,zd}\lambda_{zd}. \tag{5.5.25}$$

By following this definition, λ_{zd} can then be estimated by $\widehat{\lambda}_{zd} = \frac{N_{zd}}{N}$. Here N_{zd} is the size of the subgroup in a trial where all individuals have $Z_i = z$ and $D_i(z) = d$, and N is the total number of individuals involved in the study trial. Furthermore, we have $Pr(U_i = u|Z_i = 1) = Pr(U_i = u|Z_i = 0)$ as randomisation guarantees a balance between the treated ($Z_i = 1$) and the control ($Z_i = 0$) groups. Therefore we can write $\pi_u = Pr(U_i = u|Z_i = 1) = Pr(U_i = u|Z_i = 0)$.

Since the number of subjects who do or do not actually receive the treatment in the treatment group ($Z_i = 1$) are observable (denoted by N_{11} and N_{10} respectively), we can then estimate π_n by $\widehat{\pi}_n = \frac{N_{10}}{N_1}$. If always-takers exist in a trial, then the number of subjects N_{01} in the control group who receive the treatment is known. Then we have $\widehat{\pi}_a = \frac{N_{01}}{N_0}$. Also $\pi_c + \pi_a + \pi_n = 1$ because no defiers exist. Then the probability of compliers can be estimated by $\widehat{\pi}_c = 1 - \widehat{\pi}_a - \widehat{\pi}_n$ (O'Malley & Normand, 2005). These can be expressed as:

$$\begin{aligned} \widehat{\pi}_n &= \frac{N_{10}}{N_1}; \\ \widehat{\pi}_a &= \frac{N_{01}}{N_0}; \\ \widehat{\pi}_c &= 1 - \widehat{\pi}_a - \widehat{\pi}_n. \end{aligned} \tag{5.5.26}$$

Recall that individuals from the subgroup with $Z_i = 1$ and $D_i(1) = 1$ can be written as $i \in \mathcal{A}(1, 1)$. Their compliance type can be *compliers*, $U_i = c$, or

always-takers, $U_i = a$. Similarly, individuals from the subgroup with $Z_i = 0$ and $D_i(0) = 0$, denoted by $i \in \mathcal{A}(0,0)$, may either be *compliers*, $U_i = c$, or *never-takers*, $U_i = n$. Then, following the definition of $\pi_{u,zd}$ and by Table 5.1, we have:

$$\begin{aligned} \pi_{c,11} &= \frac{\pi_c}{\pi_c + \pi_a}, \\ \pi_{c,00} &= \frac{\pi_c}{\pi_c + \pi_n}. \end{aligned} \tag{5.5.27}$$

By following equation (5.5.14) and considering the relationships given in equation (5.5.27), the likelihood function $L^F(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$, can be replaced by a product of $l^F(\boldsymbol{\theta})$ and $l^F(\boldsymbol{\pi})$. That is:

$$L^F(\boldsymbol{\beta}) = l^F(\boldsymbol{\theta}) l^F(\boldsymbol{\pi}), \tag{5.5.28}$$

where

$$\begin{aligned} l^F(\boldsymbol{\theta}) &= \prod_{i=1} F^i_{u,zd}(\theta_{T,uz}, \theta_{C,uz}), \\ l^F(\boldsymbol{\pi}) &= \prod_{i=1} \pi_{u,zd}. \end{aligned} \tag{5.5.29}$$

Therefore, to maximise $L^F(\boldsymbol{\beta})$ is equivalent to maximise $l^F(\boldsymbol{\theta})$ and $l^F(\boldsymbol{\pi})$, separately.

Because of the relationships given equation (5.5.27), which implies that $\pi_{c,11}$ and $\pi_{c,00}$ can be calculated directly through the value of π_u for $u \in \{c, n, a\}$, we then write:

$$\boldsymbol{\pi}^* = (\pi_c, \pi_n, \pi_a). \tag{5.5.30}$$

Therefore, to estimate parameter $\boldsymbol{\beta} = (\boldsymbol{\theta}, \boldsymbol{\pi})$ is equivalent to estimating $(\boldsymbol{\theta}, \boldsymbol{\pi}^*)$. As we stated before, the MLE of $\boldsymbol{\pi}^*$ can be estimated by equation (5.5.26). Hence we need only to maximise $l^F(\boldsymbol{\theta})$ for obtaining the MLE of $\boldsymbol{\theta}$. The EM algorithm is employed here because an unobservable variable, compliance type U_i, is involved in the likelihood function $l^F(\boldsymbol{\theta})$.

In other words, there is only the first part of parameter $\boldsymbol{\beta}$, denoted by $\boldsymbol{\theta}$ need to be further estimated through the EM algorithm. That is:

$$\begin{aligned} \boldsymbol{\theta} &= (\boldsymbol{\theta}_{T,uz} \boldsymbol{\theta}_{C,uz}); \\ &= (\boldsymbol{\theta}_{T,c1}, \boldsymbol{\theta}_{T,c0}, \boldsymbol{\theta}_{T,n}, \boldsymbol{\theta}_{T,a}, \boldsymbol{\theta}_{C,c1}, \boldsymbol{\theta}_{C,c0}, \boldsymbol{\theta}_{C,n}, \boldsymbol{\theta}_{C,a}). \end{aligned} \tag{5.5.31}$$

The rest part of $\boldsymbol{\beta}$, the parameter $\boldsymbol{\pi}^*$ can be regarded as a known parameter when we estimate the parameter $\boldsymbol{\theta}$. As a consequence, the parameter $\boldsymbol{\pi}$ is also known because of equation (5.5.27). For simplification of notation, in the remains of this chapter we still use $\boldsymbol{\pi}$ rather than $\boldsymbol{\pi}^*$ to refer to the rest part of $\boldsymbol{\beta}$.

5.5.5 EM algorithm

The EM algorithm is a generally applicable algorithm that provides an iterative procedure for computing MLEs in such situations. Indeed, the EM algorithm is a commonly used method of maximum likelihood to deal with incomplete data (Laird & Louis, 1982; Baker, 1992; Meng & Rubin, 1991, 1993; Imbens & Rubin, 1997; O'Malley & Normand, 2005; Peng et al., 2004).

In general, the likelihood method requires the score equation to be solved. Unfortunately, the likelihood function of $\boldsymbol{\beta}$ for incomplete data, as given in equation (5.5.16), cannot be formulated. This is because, at least for those individuals assigned to the control group, their compliance type, U_i, will never be observed. The EM algorithm approaches this problem (of solving the incomplete data likelihood score function) indirectly by proceeding iteratively in terms of the complete data log likelihood function of $\boldsymbol{\beta}$, denoted by $log(L^F(\boldsymbol{\beta}))$ (equation (5.5.28)).

The expectation step

According to Dempster et al. (1977), as introduced in Section 5.5.2, the first step in the EM algorithm is to obtain the conditional expectation of $log(L^F(\boldsymbol{\beta}))$, the Q-function $Q(\boldsymbol{\beta}, \boldsymbol{\beta}^k)$. Now it is replaced by the conditional expectation of $log(l^F(\boldsymbol{\theta}))$, the Q-function $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^k)$. In other words:

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^k) = E_u\{log(l^F(\boldsymbol{\theta})|(T_i,\ C_i)^{obs},\ Z_i = z,\ D_i(z) = d),\ U_i = u, ; \boldsymbol{\theta}^k\}, \tag{5.5.32}$$

where $u \in \{c, n, a\}$ and $l^F(\boldsymbol{\theta})$ denotes the likelihood function with respect to parameter $\boldsymbol{\theta}$ for the complete data that is given in equation (5.5.29). $(T_i,\ C_i)^{obs}$ indicates the observable part of survival outcomes $(T_i,\ C_i)$, which is $(T_i^{obs} = t,\ C_i^{obs} > t)$ if the *ith* individual experienced the event at time t;

or $(T_i^{obs} > t,\ C_i^{obs} = t)$ if the *ith* individual is censored at time t.

Note that equations (5.5.27) and (5.5.26) have shown that $\boldsymbol{\pi}$ is only determined by the observed values of subgroup sizes, which are denoted by N_1, N_{10}, N_0, and N_{01}. This implies that for a certain dataset, $L^F(\boldsymbol{\pi})$ would be a constant because in the dataset N_1, N_{10}, N_0, and N_{01} are fixed. Hence, to maximise $l^F(\boldsymbol{\theta})$ is equivalent to maximise $L^F(\boldsymbol{\beta}|\boldsymbol{\pi}) = l^F(\boldsymbol{\theta})l^F(\boldsymbol{\pi}) = Kl^F(\boldsymbol{\theta})$ when $\boldsymbol{\pi}$ is known. (K refers to any constant here). So we give the likelihood function with respect to parameter $\boldsymbol{\theta}$ as $l^F(\boldsymbol{\theta}) = L^F(\boldsymbol{\beta}|\boldsymbol{\pi})$, which has the same structure as $L^F(\boldsymbol{\beta})$, the likelihood function of $\boldsymbol{\beta}$ given in equation (5.5.15), but $\boldsymbol{\pi}$ is viewed as a known vector. The advantage of using this form of the likelihood function $l^F(\boldsymbol{\theta}) = L^F(\boldsymbol{\beta}|\boldsymbol{\pi})$ will be seen in Chapter 7 when we create an extended PPO-survival model for a more complicated situation in time-to-event study.

As compliance type, U_i, can be treated as a discrete random variable with $c,\ a,\ n$ (three states), then, by the definition of expectation of discrete random variables, equation (5.5.32) can be expressed as follows:

$$\begin{aligned} &Q^i(\boldsymbol{\theta},\boldsymbol{\theta}^k) \\ &\quad = \sum_{u\in\{c,n,a\}} \left\{ \begin{array}{l} Pr(U_i = u|(T_i,\ C_i)^{obs},\ Z_i = z,\ D_i(z) = d;\ \boldsymbol{\theta}^k) \\ \{log(l(\boldsymbol{\theta}, u)\} \end{array} \right\} \end{aligned} \tag{5.5.33}$$

where $l(\boldsymbol{\theta}, u)$ is another form of $l^F(\boldsymbol{\theta})$, the likelihood function with respect to θ for complete data, we give its form as follows:

$$l^F(\boldsymbol{\theta}) = L(\boldsymbol{\beta}, u|\boldsymbol{\pi}), \tag{5.5.34}$$

because $L^F(\boldsymbol{\beta}) = L(\boldsymbol{\beta}, u)$.

In addition, we can further assume that compliance type, U_i, follows the Bernoulli distribution, given observed data $T_i(z),\ C_i(z),\ D_i(z)$. This is true because of the assumptions of *compound exclusion restriction* and *monotonicity*, given $(Z_i = z,\ D_i(z) = d)$. Recall that compliance type has up to two compliance categories for this development (refer to Table 5.1). More details about the structure of the Q-function $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^k)$, the expectation function, are given Section 5.9.2.

The maximisation step

After obtaining the formulation of Q-function $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^k)$ given the current parameter values $\boldsymbol{\theta}^k$, the second step of the EM algorithm is to find a new estimate, $\boldsymbol{\theta}^{k+1}$, which maximizes the Q-function $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^k)$. This implies that the following inequality is always satisfied:

$$Q(\boldsymbol{\theta}^{k+1}, \boldsymbol{\theta}^k) \geq Q(\boldsymbol{\theta}, \boldsymbol{\theta}^k) \quad \text{for all } \theta \in \Omega_\theta, \tag{5.5.35}$$

where $\Omega_{\boldsymbol{\theta}}$ is the parameter space of $\boldsymbol{\theta}$.

Further information on the process of maximisation for the non-ignorable PPO-survival model is given in Section 5.9.3.

5.6 Model specifications

For the purpose of showing how our non-ignorable models work, we focus mainly on the Weibull model as the survival distribution. This is due to its special features. Firstly, the Weibull distribution is flexible for many different patterns of data. This can be seen from its varying types of hazard curve (Klein, 1997; Cox & Oakes, 1984). Secondly, the Weibull distribution is associated with relatively simple survival, hazards and probability density functions (equations (5.6.1), (5.6.2) and (5.6.3)). This makes the Weibull a very popular parametric model in conventional survival analysis. Furthermore, the Weibull distribution is the only distribution which satisfies both the accelerated failure time (AFT) and the Cox proportional hazard (PH) models (Cox & Oakes, 1984). The log-normal model is another plausible parametric distribution used in conventional survival analysis. We also specify the log-normal distribution for the non-ignorable PPO-survival model (Section 5.6.3). In this thesis, our parametric potential-outcome (PPO) models are specified in terms of the Weibull and the log-normal distributions. Other distributions are the topic of future research.

5.6.1 Weibull distribution

Following standard survival analytic procedures, we suppose that the time-to-event of interest (or failure time) has a Weibull distribution (Klein, 1997;

Cox & Oakes, 1984; Kalbfleisch & Prentice, 2002). The Weibull distribution, which has a variable shape parameter, can be conveniently used to describe many different patterns of survival data. It is well known that the hazard function in a Weibull model varies according to the value of its shape parameter. It may be increasing, decreasing, constant, or possess some other characteristics which describe the failure mechanism (Klein, 1997; Cox & Oakes, 1984; Kalbfleisch & Prentice, 2002).

The Weibull failure time distribution (T) is parameterized by (ρ, α), where ρ and α represent scale and shape parameters, respectively. So the *density function* and *survival function* corresponding to a Weibull distribution can be expressed as follows:

$$f_w(t) = \alpha\rho(\rho t)^{(\alpha-1)} \exp(-(\rho t)^{\alpha}), \tag{5.6.1}$$

$$S_w(t) = \exp(-(\rho t)^{\alpha}); \tag{5.6.2}$$

and the *hazard rate function* can be written as:

$$h_w(t) = \alpha\rho(\rho t)^{(\alpha-1)}, \tag{5.6.3}$$

(p. 19, Cox & Oakes (1984)).

5.6.2 Log-normal distribution

Since in time-to-event studies, the outcome of interest, T (the event time), is almost certainly a positive value ($T > 0$), the most common way of realising the relevant data transformation is to take the natural logarithm of event time, $\log(T)$. If the transformed data $\log(T)$ follow a normal distribution, by definition, the event time T will possess a log-normal distribution. Like the normal distribution, the log-normal distribution is completely specified by two parameters, μ and σ, the mean and variance of $log(T)$, respectively. Its *density function* is expressed as follows:

$$f_l(t) = \frac{\exp[-\frac{1}{2}(\frac{\log[t]-\mu}{\sigma})^2]}{t(2\pi)^{1/2}\sigma} = \phi(\frac{\log[t]-\mu}{\sigma})/t, \tag{5.6.4}$$

where $\phi()$ represents the standard normal probability density function, which can be written as $\phi(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$ (p. 39, Klein (1997)).

By following the definition of the *survival function*, it is straightforward to obtain its form for the log-normal model, namely:

$$S_l(t) = 1 - \Phi(\frac{\log[t] - \mu}{\sigma}), \tag{5.6.5}$$

where $\Phi()$ is the cumulative distribution function of a standard normal variable, which can be expressed as $\Phi(x) = \int_{-\infty}^{x} \phi(u)du$ (p. 39, Klein (1997)). The *hazard function* in the log-normal model can be expressed as a function of its density and survival function, namely:

$$h_l(t) = \frac{f_l(t)}{S_l(t)}, \tag{5.6.6}$$

(p. 37, Klein (1997)).

The hazard function of the log-normal is hump-shaped. Specifically, its value at $t = 0$ is zero, $h_l(0) = 0$, and it increases to a maximum and then decreases to zero when the time variable t approaches infinity ($t \to \infty$). Note that the *hazard function* decreases for large values of t, which seems implausible in many time-to-event studies. This is precisely why the log-normal model has been criticised when chosen as a lifetime distribution. The log-normal distribution may fit certain cases, however, where large values of t are not of interest (Klein, 1997).

5.6.3 Specific non-ignorable PPO-survival models

When accounting for a censoring time distribution, the non-ignorable survival model can be specified in more than two forms. This is true, even when the Weibull and the log-normal are the only pre-specified distributions. Note that failure and censoring times may follow different types of distribution. For instance, suppose that the failure time follows a Weibull distribution and its associated censoring time has an underlying log-normal distribution. In this case, the non-ignorable survival model has to be specified as a NIGN-WL model; namely a non-ignorable survival Weibull log-normal model. However, when failure and censoring times follow the log-normal and

the Weibull distribution, respectively, we have the NIGN-LW model. In this thesis, we focus, however, on specific non-ignorable survival models with the same distributions for both the failure and censoring times. These are clearly denoted as the NIGN-WW model and the NIGN-LL model. Table 5.2 gives the nomenclature for the non-ignorable survival model specifications.

Table 5.2: Specification of the non-ignorable PPO-survival models.

Distribution	non-ignorable survival model
Weibull	NIGN-WW model
log-normal	NIGN-LL model

5.7 The AFT model

To simplify the model structure, especially to simplify model parameter β, we adopt an accelerated failure time (AFT) model approach (Cox & Oakes, 1984; Klein, 1997). In this section, we first prove a useful property of AFT for the Weibull or log-normal distribution (Section 5.7.1), then focus on the formulation and application for the simplest case, where no pre-treatment covariate is involved (Section 5.7.2 and Section 5.7.3). The complicated form for incorporating pretreatment covariates is, however, given in Section 7.4.

5.7.1 An AFT property with its proof

Theorem 1. *Suppose a failure time variable T satisfies an accelerated failure time (AFT) model $T = T_0 \exp(\beta' X)$, where T_0 represents the failure time corresponding to the baseline; X and β denote covariate vector and its associated regression coefficients, both of them are column vectors with the same length p.*

1. *If $T_0 = t_0 \sim Weib(\rho_0, \alpha_0)$, t_0 follows a Weibull distribution with scale parameter ρ_0 and shape parameter α_0, the new variable $T = t$ then follows a Weibull distribution with $\rho = \rho_0 \exp(-\beta' X)$ and $\alpha = \alpha_0$, namely $t \sim Weib(\rho, \alpha)$.*

2. *If $T_0 = t_0 \sim Logn(\mu_0, \sigma_0)$, t_0 follows a log-normal distribution with two parameters, μ_0 and σ_0, the mean and variance of $log(T_0)$, the new variable $T = t$ then follows a log-normal distribution with $\mu = \mu_0 + \beta X'$ and $\sigma = \sigma_0$, i.e. $T = t \sim Logn(\mu, \sigma)$.*

Proof. By following the definition of the survival function with respect to the failure time T, the survival function can be expressed as follows:

$$\begin{aligned} S(t) = Pr(T > t) &= Pr(T_0 \exp(\beta' X) > t); \quad \because T = T_0 \exp(\beta' X); \\ &= Pr(T_0 > t \exp(-\beta' X)); \quad \because \exp(\beta' X) > 0; \\ &= S_0(t \exp(-\beta' X)); \quad \text{by the definition of } S_0(t). \end{aligned} \tag{5.7.1}$$

In addition, by the definition of a probability density function (p.d.f.), we have the following:

$$\begin{aligned} f(t) &= -\frac{dS(t)}{dt}; \quad \text{(by definition)}; \\ &= \frac{dS_0(t_0)}{dt_0}\frac{dt_0}{dt}; \quad \text{for } t_0 = t \exp(-\beta' X) \\ &= f_0(t_0) \exp(-\beta' X). \end{aligned} \tag{5.7.2}$$

1) $\because T_0 = t_0 \sim \text{Weib}(\rho_0, \alpha_0)$, we have

$$f_{w0}(t) = \alpha_0 \rho_0 (\rho_0 t)^{(\alpha_0 - 1)} \exp(-(\rho_0 t)^{\alpha_0}), \tag{5.7.3}$$

$$S_{w0}(t) = \exp(-(\rho_0 t)^{\alpha_0}). \tag{5.7.4}$$

where $f_{w0}(t)$ and $S_{w0}(t)$ denote the probability density (p.d.f.) and survival functions (S) with respect to T_0.

By following equations (5.7.1) and (5.7.2), the survival function and the probability density function (p.d.f.), $f_w(t)$, with respect to the failure time T in a Weibull model, can be expressed as:

$$\begin{aligned} S_w(t) = Pr(T > t) &= S_{w0}(t \exp(-\beta' X)); \\ &= \exp(-[\rho_0 t \exp(-\beta' X)]^{\alpha_0}), \end{aligned} \tag{5.7.5}$$

and the probability density function (p.d.f.), $f_w(t)$ as:

$$\begin{aligned} f_w(t) &= f_{w0}(t\exp(-\beta' X))\exp(-\beta' X); \\ &= \alpha_0\rho_0(\rho_0 t\exp(-\beta' X))^{(\alpha_0-1)}\exp(-[\rho_0 t\exp(-\beta' X)]^{\alpha_0})\exp(-\beta' X); \\ &= \alpha_0\rho_0\exp(-\beta' X)(\rho_0 t\exp(-\beta' X))^{(\alpha_0-1)}\exp(-[\rho_0 t\exp(-\beta' X)]^{\alpha_0}). \end{aligned} \tag{5.7.6}$$

Let $\rho = \rho_0\exp(-\beta' X)$ and $\alpha = \alpha_0$. Equations (5.7.7) and (5.7.7) can then be re-written as:

$$\begin{aligned} S_w(t) = Pr(T > t) &= \exp(-[\rho_0 t\exp(-\beta' X)]^{\alpha_0}) \\ &= \exp(-(\rho t)^{\alpha}), \end{aligned} \tag{5.7.7}$$

and

$$\begin{aligned} f_w(t) = &= \alpha_0\rho_0\exp(-\beta' X)(\rho_0\exp(-\beta' X)t)^{(\alpha_0-1)}\exp(-[\rho_0\exp(-\beta' X)t]^{\alpha_0}); \\ &= \alpha\rho(\rho t)^{(\alpha-1)}\exp(-(\rho t)^{\alpha}). \end{aligned} \tag{5.7.8}$$

Because equations (5.7.7) or (5.7.8) hold, we have $T = t \sim \text{Weib}(\rho, \alpha)$ where $\rho = \rho_0\exp(-\beta' X)$ and $\alpha = \alpha_0$.

2) $\because T_0 = t_0 \sim Logn(\mu_o, \sigma_0)$, we can write:

$$f_{l0}(t_0) = \frac{\exp[-\frac{1}{2}(\frac{\log[t_0]-\mu_0}{\sigma_0})^2]}{t_0(2\pi)^{1/2}\sigma_0} = \phi(\frac{\log[t_0]-\mu}{\sigma})/t_0; \tag{5.7.9}$$

where $\phi()$ represents the standard normal probability density function, which can be written as $\phi(x) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{x^2}{2})$; and

$$S_{l0}(t) = 1 - \Phi(\frac{\log[t]-\mu_0}{\sigma_0}); \tag{5.7.10}$$

where $\Phi()$ is the cumulative distribution function of a standard normal variable, which has the general form $\Phi(x) = \int_{-\infty}^{x}\phi(u)du$. By following equations (5.7.1) and (5.7.2), the survival function and the probability density function (p.d.f.) $f_l(t)$ with respect the failure time T in a log-normal model can be expressed as follows:

$$
\begin{aligned}
S_l(t) = Pr(T > t) = S_{l0}(t\exp(-\beta' X)); \\
&= 1 - \Phi(\frac{\log[t\exp(-\beta' X)] - \mu_0}{\sigma_0}); \\
&1 - \Phi(\frac{\log[t] - \beta' X - \mu_0}{\sigma_0}); \qquad (5.7.11)
\end{aligned}
$$

and

$$
\begin{aligned}
f_l(t) &= f_{l0}(t\exp(-\beta' X))\exp(-\beta' X); \\
&= \frac{\exp[-\frac{1}{2}(\frac{\log[t\exp(-\beta' X)]-\mu_0}{\sigma_0})^2]}{t\exp(-\beta' X)(2\pi)^{1/2}\sigma_0}\exp(-\beta' X); \\
&= \frac{\exp[-\frac{1}{2}(\frac{\log[t]-\beta' X-\mu_0}{\sigma_0})^2]}{t\exp(-\beta' X)(2\pi)^{1/2}\sigma_0}\exp(-\beta' X); \\
&= \frac{\exp[-\frac{1}{2}(\frac{\log[t]-(\beta' X+\mu_0)}{\sigma_0})^2]}{t(2\pi)^{1/2}\sigma_0}. \qquad (5.7.12)
\end{aligned}
$$

Using $\mu = \mu_0 + \beta' X$ and $\sigma = \sigma_0$, equations (5.7.11) and (5.7.12) can be re-written as follows:

$$
S_l(t) = 1 - \Phi(\frac{\log[t] - \mu}{\sigma}), \qquad (5.7.13)
$$

and

$$
f_l(t) = \frac{\exp[-\frac{1}{2}(\frac{\log[t]-\mu}{\sigma})^2]}{t(2\pi)^{1/2}\sigma}. \qquad (5.7.14)
$$

Therefore, $T = T_0\exp(\beta' X)$ follows a log-normal distribution with the parameters μ and σ. Explicitly, $T = t \sim Logn(\mu, \sigma)$ where $\mu = \mu_0 + \beta' X$ and $\sigma = \sigma_0$. □

5.7.2 The AFT model for failure time

In general, the simplest form of the AFT model can be written as $T_i(1) = T_i(0)/\psi$, where ψ is often given as an exponential function $\psi = \exp(\beta^T \mathbf{Z})$ where $\mathbf{Z}$ denotes covariates and β is a vector parameter that is associated with those related covariates variables (p.64, Cox & Oakes (1984)) and is the

so-called *accelerated factor*(p.47, Klein (1997)).

Being based on the potential outcome framework, the AFT model can be constructed in two parts, the first modelling the relationship between potential outcomes (the causal model) and the second modelling differences between compliance types (selection model). Thus we first assume the simplest form as the following, where compliance type is presumed to have only two categories:

$$\begin{aligned} T_i(1) &= T_i(0)\exp(\zeta + \zeta_U U_i); \\ T_i(0) &= T_{0i}\exp(\phi_U U_i); \\ T_{0i} &\sim D(\theta_T^{(1)}, \theta_T^{(2)}), \end{aligned} \tag{5.7.15}$$

where T_{0i} denotes failure time of baseline and $D(\theta_T^{(1)}, \theta_T^{(2)})$ represents a standard distribution, such as the Weibull distribution, and $(\theta_T^{(1)}, \theta_T^{(2)})$ denotes its associated parameters. For example, if the standard distribution is given by a Weibull distribution, then $\theta_T^{(1)} = \rho_T^0$ and $\theta_T^{(2)} = \alpha_T^0$. In this case, we have $D(\theta_T^{(1)}, \theta_T^{(2)}) \simeq W(\rho_T^0, \rho_T^0)$.

However, the *ith* individual's compliance type, $U_i = u$, can possibly fall into one of three categories: *complier*, *always-taker* and *never-taker* ($u \in \{c, a, n\}$). So we can use two dummy variables, i.e. $U_i^{(c)}$ and $U_i^{(a)}$, to denote U_i, that is $U_i = (U_i^{(c)}, U_i^{(a)})^T$. More specifically, $(U_i^{(c)}, U_i^{(a)})^T = (1, 0)^T$ is equivalent to $U_i = c$, which indicates a *complier*; $(U_i^{(c)}, U_i^{(a)})^T = (0, 1)^T$ is equivalent to $U_i = a$, which indicates an *always-taker*, and $(U_i^{(c)}, U_i^{(a)})^T = (0, 0)^T$ is equivalent to $U_i = n$, which indicates a *never-taker*. Table 5.3 gives the relationship between dummy variables and the categorical variable for compliance type.

Table 5.3: Dummy variables for compliance type U_i.

$U_i = u$	$(U_i^{(c)}$	$U_i^{(a)})$
c	(1	0)
a	(0	1)
n	(0	0)

Let $\zeta_U = (\zeta^{(c)}, \zeta^{(a)})$, and $\phi_U = (\phi^{(c)}, \phi^{(a)})$, then equation (5.7.15) can be

re-written as follows:

$$\begin{aligned} T_i(1) &= T_i(0)\exp(\zeta + \zeta^{(c)}U_i^{(c)} + \zeta^{(a)}U_i^{(a)}); \\ T_i(0) &= T_{0i}\exp(\phi^{(c)}U_i^{(c)} + \phi^{(a)}U_i^{(a)}); \\ T_{0i} &\sim D(\theta_T^{(1)}, \theta_T^{(2)}). \end{aligned} \tag{5.7.16}$$

The first equality in equation (5.7.16) is a model for the relationship between the two potential outcomes. This model represents the effect of assigned treatment, which is assumed to multiply the failure time that would be observable if not assigned to treatment ($T_i(0)$) by $\exp(\zeta)$ for *never-takers* ($(U_i^{(a)}, U_i^{(c)}) = (0,0)$), by $\exp(\zeta + \zeta^{(a)})$ for *always-takers* ($(U_i^{(a)}, U_i^{(c)}) = (1,0)$) or by $\exp(\zeta+\zeta^{(c)})$ for *compliers* ($(U_i^{(a)}, U_i^{(c)}) = (0,1)$). These can be written as:

$$T_i(1) = \begin{cases} T_i(0)\exp(\zeta + \zeta^{(c)}) & \text{for } U_i = c; \\ T_i(0)\exp(\zeta + \zeta^{(a)}) & \text{for } U_i = a; \\ T_i(0)\exp(\zeta) & \text{for } U_i = n. \end{cases} \tag{5.7.17}$$

In other words, the effect of assigned treatment is allowed to vary with compliance type. However, by following the *exclusion restriction* assumption, which implies that $T_i(1) = T_i(0)$ for $U_i = u$ and $u \in \{a, n\}$, we have the following equations:

$$\begin{aligned} \exp(\zeta) &= 1; \\ \exp(\zeta + \zeta^{(a)}) &= 1; \end{aligned} \tag{5.7.18}$$

which are equivalent to the following constraints:

$$\begin{aligned} \zeta &= 0; \\ \zeta^{(a)} &= 0; \end{aligned} \tag{5.7.19}$$

This means that under the *exclusion restriction* assumption, these constraints, $\zeta = \zeta^{(a)} = 0$, should be imposed because under this restriction, it is quite reasonable to assume that, since assigned treatment does not change the treatment received for *always-takers* and *never-takers*, there is no causal effect of assigned treatment for these groups.

The second equality in equation (5.7.16) models the effect of compliance

type, if not assigned to treatment. Under this model, the failure time, if not assigned to the treatment, is T_{0i} for *never-takers*, $T_{0i}\exp(\phi^{(a)})$ for *always-takers* and $T_{0i}\exp(\phi^{(c)})$ for *compliers*. In other words, we have:

$$T_i(0) = \begin{cases} T_{0i}\exp(\phi^{(c)}) & \text{for } U_i = c, \\ T_{0i}\exp(\phi^{(a)}) & \text{for } U_i = a, \\ T_{0i} & \text{for } U_i = n. \end{cases} \tag{5.7.20}$$

The two expressions given in equation (5.7.16) imply that the model for the potential outcome corresponding to being assigned to treatment $Z_i = 1$, $T_i(1)$ can be written as follows:

$$T_i(1) = T_{0i}\exp(\zeta + (\zeta^{(c)} + \phi^{(c)})U_i^{(c)} + (\zeta^{(a)} + \phi^{(a)})U_i^{(a)}); \tag{5.7.21}$$

which is reduced as

$$T_i(1) = T_{0i}\exp((\zeta^{(c)} + \phi^{(c)})U_i^{(c)} + \phi^{(a)}U_i^{(a)}); \tag{5.7.22}$$

under the exclusion restriction $\zeta = \zeta^{(a)} = 0$.

Note that the AFT model makes quite strong assumptions: it effectively models individual-level causal effects. For example, the model assumes $T_i(1)/T_i(0) = \zeta^{(c)}$ for all compliers.

The relationships between the categorial variable U_i and the dummy variables ($U_i^{(c)}$, $U_i^{(a)}$), which were given in Table 5.3, have been used in equations (5.7.17) and (5.7.20). Given the values of the vector $(U_i^{(c)}, U_i^{(a)})$ which corresponds to the three different compliance type U_i (Table 5.3), equation (5.7.16) can then be re-written:

$$T_i(1) = \begin{cases} T_{0i}\exp(\zeta^{(c)} + \phi^{(c)}) & \text{for } U_i = c; \\ T_{0i}\exp(\phi^{(a)}) & \text{for } U_i = a; \\ T_{0i} & \text{for } U_i = n; \end{cases} \tag{5.7.23}$$

and

$$T_i(0) = \begin{cases} T_{0i}\exp(\phi^{(c)}) & \text{for } U_i = c, \\ T_{0i}\exp(\phi^{(a)}) & \text{for } U_i = a, \\ T_{0i} & \text{for } U_i = n; \end{cases} \tag{5.7.24}$$

where $T_{0i} \sim D(\theta_T^{(1)}, \theta_T^{(2)})$.

5.7.3 The AFT model for censoring time

Similarly, for censoring time, the AFT model based on the potential outcomes framework can be expressed as follows:

$$\begin{aligned} C_i(1) &= C_i(0)\exp(\xi + \xi_U U_i), \\ C_i(0) &= C_{0i}\exp(\varphi_U U_i), \\ C_{0i} &\sim D(\theta_C^{(1)}, \theta_C^{(2)}); \end{aligned} \tag{5.7.25}$$

where C_{0i} denotes censoring time of baseline, and $D(\theta_C^{(1)}, \theta_C^{(2)})$ represents a standard distribution, and $\theta_C^{(1)}, \theta_C^{(2)}$ denote its associated parameters. For example, if the standard distribution is given as a Weibull distribution, then $\theta_C^{(1)} = \rho_C^0$ and $\theta_C^{(2)} = \alpha_C^0$.

By following the logic used in Section 5.7.2, we have the simplified AFT model for censoring time, that is:

$$C_i(1) = C_{0i}\exp((\xi^{(c)} + \varphi^{(c)})U_i^{(c)} + \varphi^{(a)}U_i^{(a)}); \tag{5.7.26}$$

under the exclusion restriction $\xi = \xi^{(a)} = 0$.

Now equation (5.7.25) is equivalent to:

$$C_i(1) = \begin{cases} C_{0i}\exp(\xi^{(c)} + \varphi^{(c)}) & \text{for } U_i = c, \\ C_{0i}\exp(\varphi^{(a)}) & \text{for } U_i = a, \\ C_{0i} & \text{for } U_i = n; \end{cases} \tag{5.7.27}$$

and

$$C_i(0) = \begin{cases} C_{0i}\exp(\varphi^{(c)}) & \text{for } U_i = c, \\ C_{0i}\exp(\varphi^{(a)}) & \text{for } U_i = a, \\ C_{0i} & \text{for } U_i = n; \end{cases} \tag{5.7.28}$$

where $C_{0i} \sim D(\theta_C^{(1)}, \theta_C^{(2)})$.

5.8 Parameter of interest $\boldsymbol{\theta}$ in the specific models

The parameter $\boldsymbol{\theta} = (\boldsymbol{\theta}_{T,uz}, \boldsymbol{\theta}_{C,uz})$ is associated with failure and censoring time distributions, respectively, corresponding to four subgroups each. In total, $\boldsymbol{\theta}$ includes $8 \times 2 = 16$ components. This is true, as both a Weibull and a log-normal distribution include two model parameters: the scale parameter ρ and the shape parameter α for the Weibull; the mean μ and the variance σ of the logarithms of survival outcome for the log-normal.

5.8.1 Parameter $\boldsymbol{\theta}$ in the NIGN-WW model

Suppose failure time T_{0i}, corresponding to the baseline, follows a Weibull distribution with parameter ρ_T^0 and α_T^0. By following Theorem 1 (proved in Section 5.7.1), the parameter of most interest, $\boldsymbol{\theta}_T = \boldsymbol{\theta}_{T,uz}$, which is associated with the distributions of failure time and is specified in equation (5.5.31) (the first four parameters corresponding to 4 possible subgroups), has the following relationships with the parameters associated with the standard distribution for the baseline failure time, ρ_T^0 and α_T^0, and the regression coefficients in AFT model, which are:

$$\begin{aligned}
&\boldsymbol{\theta}_{T,c1} = (\rho_{T,c1}, \alpha_{T,c1}); && \text{where } \rho_{T,c1} = \rho_T^0 \exp(\zeta^{(c)} + \phi^{(c)}) \text{ and } \alpha_{T,c1} = \alpha_T^0,\\
&\boldsymbol{\theta}_{T,c0} = (\rho_{T,c0}, \alpha_{T,c0}); && \text{where } \rho_{T,c0} = \rho_T^0 \exp(\phi^{(c)}) \text{ and } \alpha_{T,c0} = \alpha_T^0,\\
&\boldsymbol{\theta}_{T,n} = (\rho_{T,n}, \alpha_{T,n}); && \text{where } \rho_{T,n} = \rho_T^0 \text{ and } \alpha_{T,n} = \alpha_T^0,\\
&\boldsymbol{\theta}_{T,a} = (\rho_{T,a}, \alpha_{T,a}); && \text{where } \rho_{T,a} = \rho_T^0 \exp(\phi^{(a)}) \text{ and } \alpha_{T,c0} = \alpha_T^0.
\end{aligned} \tag{5.8.1}$$

Similarly, we suppose that the censoring time C_0, corresponding to the baseline, also follows a Weibull distribution with parameters ρ_C^0 and α_C^0. In this case, $\boldsymbol{\theta}_C = \boldsymbol{\theta}_{C,uz}$) can be specified as follows:

$$\begin{aligned}
&\boldsymbol{\theta}_{C,c1} = (\rho_{C,c1}, \alpha_{C,c1}); && \text{where } \rho_{C,c1} = \rho_C^0 \exp(\xi^{(c)} + \varphi^{(c)}) \text{ and } \alpha_{C,c1} = \alpha_C^0,\\
&\boldsymbol{\theta}_{C,c0} = (\rho_{C,c0}, \alpha_{C,c0}); && \text{where } \rho_{C,c0} = \rho_C^0 \exp(\varphi^{(c)}) \text{ and } \alpha_{C,c0} = \alpha_C^0,\\
&\boldsymbol{\theta}_{C,n} = (\rho_{C,n}, \alpha_{C,n}); && \text{where } \rho_{C,n} = \rho_C^0 \text{ and } \alpha_{C,n} = \alpha_C^0,\\
&\boldsymbol{\theta}_{C,a} = (\rho_{c,a}, \alpha_{C,a}); && \text{where } \rho_{C,a} = \rho_C^0 \exp(\varphi^{(a)}) \text{ and } \alpha_{C,c0} = \alpha_C^0.
\end{aligned} \tag{5.8.2}$$

Consequently, the parameter $\boldsymbol{\theta}$ given in equations (5.5.31) for the NIGN-

WW model can be written as follows:

$$\begin{aligned}\boldsymbol{\theta} &= (\boldsymbol{\theta}_{T,uz}, \boldsymbol{\theta}_{C,uz}),\\ &= (\rho_{T,c1}, \rho_{T,c0}, \rho_{T,n}, \rho_{T,a}, \alpha_T, \rho_{C,c1}, \rho_{C,c0}, \rho_{C,n}, \rho_{C,a}, \alpha_C).\end{aligned} \tag{5.8.3}$$

Here $\alpha_T = \alpha_T^0$ and $\alpha_C = \alpha_C^0$ denote the shape parameters of the Weibull models associated with the failure and censoring times (which correspond to the baseline), respectively. The number of components involved in θ is now reduced from 16 to 10. Note that equations (5.8.1) and (5.8.2) show that these scale parameters (corresponding to the four subgroups), namely $\rho_{T,c1}, \rho_{T,c0}, \rho_{T,n}, \rho_{T,a}$, have been linked to ρ_T^0 (the scale parameter in the Weibull distribution, associated with the failure time for the baseline), through $\zeta^{(c)}, \phi^{(c)}, \phi^{(a)}$, the regression coefficients in the AFT model. However, we use the scale parameters associated with the four subgroups, rather than the AFT regression coefficients because at this stage, the former clearly describe model parameters in which each subgroup (either for failure time or censoring time) has its own scale parameter and these subgroups share the same shape parameter. Recall that these four subgroups are as follows: Subgroup 1 and Subgroup 2 consist of compliers who are all assigned to the treated group ($U_i = c$ and $Z_i = 1$) or the control group ($U_i = c$ and $Z_i = 0$), respectively; Subgroup 3 includes never-takers ($U_i = a$) only; and Subgroup 4 includes always-takers ($U_i = a$).

5.8.2 Parameter $\boldsymbol{\theta}$ in the NIGN-LL model

As in the NIGN-WW model, the parameter $\boldsymbol{\theta}$ given in equation (5.5.31) for the NIGN-LL model can then be specified as follows:

$$\begin{aligned}\boldsymbol{\theta} &= (\boldsymbol{\theta}_{T,uz}, \boldsymbol{\theta}_{C,uz}),\\ &= (\mu_{T,c1}, \mu_{T,c0}, \mu_{T,n}, \mu_{T,a}, \sigma_T, \mu_{C,c1}, \mu_{C,c0}, \mu_{C,n}, \mu_{C,a}, \sigma_C).\end{aligned} \tag{5.8.4}$$

The parameters σ_T and σ_C, given in equations (5.8.4), denote the variance parameters of the logarithm of the failure time $log(T_{0i})$ ($\theta_T^{(2)}$) and the logarithm of the censoring times $log(C_{0i})$ ($\theta_C^{(2)}$) associated with the baseline $\sigma_T = \sigma_T^0$ and $\sigma_C = \sigma_C^0$ (which were given in equations (5.7.16) and (5.7.25)).

By following Theorem 1, given in Section 5.7.1, the parameters associated with the failure time distribution satisfy the following equations:

$$\begin{aligned}
&\boldsymbol{\theta}_{T,c1} = (\mu_{T,c1}, \sigma_{T,c1}) && \text{where } \mu_{T,c1} = \mu_T^0 + \zeta^{(c)} + \phi^{(c)} \text{ and } \sigma_{T,c1} = \sigma_T^0,\\
&\boldsymbol{\theta}_{T,c0} = (\mu_{T,c0}, \sigma_{T,c0}) && \text{where } \mu_{T,c0} = \mu_T^0 + \phi^{(c)} \text{ and } \sigma_{T,c0} = \sigma_T^0,\\
&\boldsymbol{\theta}_{T,n} = (\mu_{T,n}, \sigma_{T,n}) && \text{where } \mu_{T,n} = \rho_T^0 \text{ and } \sigma_{T,n} = \sigma_T^0,\\
&\boldsymbol{\theta}_{T,a} = (\mu_{T,a}, \sigma_{T,a}) && \text{where } \mu_{T,a} = \rho_T^0 + \phi^{(a)} \text{ and } \sigma_{T,c0} = \sigma_T^0.
\end{aligned} \tag{5.8.5}$$

Parameters associated with the censoring time distribution have the following relationships:

$$\begin{aligned}
&\boldsymbol{\theta}_{C,c1} = (\mu_{C,c1}, \sigma_{C,c1}) && \text{where } \mu_{C,c1} = \mu_0^C + \xi^{(c)} + \varphi^{(c)} \text{ and } \sigma_{C,c1} = \sigma_C^0,\\
&\boldsymbol{\theta}_{C,c0} = (\mu_{C,c0}, \sigma_{C,c0}) && \text{where } \rho_{C,c0} = \mu_0^C + \varphi^{(c)} \text{ and } \sigma_{C,c0} = \sigma_C^0,\\
&\boldsymbol{\theta}_{C,n} = (\mu_{v,n}, \sigma_{C,n}) && \text{where } \rho_{C,n} = \rho_C^0 \text{ and } \sigma_{C,n} = \sigma_C^0,\\
&\boldsymbol{\theta}_{C,a} = (\mu_{C,a}, \sigma_{C,a}) && \text{where } \rho_{C,a} = \rho_C^0 + \varphi^{(a)} \text{ and } \sigma_{C,c0} = \sigma_C^0.
\end{aligned} \tag{5.8.6}$$

Table 5.4 shows the parameters involved in the NIGN-WW and the NIGN-LL models.

Table 5.4: Non-ignorable PPO-survival model parameter vectors.

Model name	model parameter vector
NIGN-WW	$\rho_{T,c1},\ \rho_{T,c0},\ \rho_{T,n},\ \rho_{T,a},\ \alpha_T,\ \rho_{C,c1},\ \rho_{C,c0}, \rho_{C,n},\ \rho_{C,a},\ \alpha_C$
NIGN-LL	$\mu_{T,c1},\ \mu_{T,c0},\ \mu_{T,n},\ \mu_{T,a},\ \sigma_T,\ \mu_{C,c1},\ \mu_{C,c0}, \mu_{C,n},\ \mu_{C,a},\ \sigma_C$

5.9 Computational issues

Since the compliance type variable, $U_i = u$, is unobservable, those general MLE methods are not easily implemented for the PPO-survival model. Alternatively, as mentioned before (Section 5.5.5), the EM algorithm (Dempster et al., 1977) can be employed for our developments. Since the posterior probabilities of compliance type U_i, $Pr(U_i = u|(T_i,\ C_i)^{obs}, Z_i = z, D_i(z) = d;\ \boldsymbol{\theta}^k)$, have been involved in the Q-function, $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^k)$, (equation (5.5.33)), we need to formulate it first before giving the specified Q-function $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^k)$.

5.9.1 Posterior probabilities of compliance type U_i

For convenience of expression, we use P^i_{ac}, P^i_{nc}, P^i_n and P^i_a to denote the related posterior probabilities for non-censored $R_i = 1$ associated with the following cases: (1) $Z_i = 1$, $D_i(1) = 1$; (2) $Z_i = 0$, $D_i(0) = 0$; (3) $Z_i = 1$, $D_i(1) = 0$; and (4) $Z_i = 0$, $D_i(0) = 1$. Note that in the first two cases given above, compliance type U_i can possible be two categories, say *compliers* ($U_i = c$) or *always-takers* ($U_i = a$) for case (1) and *compliers* ($U_i = c$) or *never-takers* ($U_i = n$) for case (2); while in case (3), compliance type U_i can only be *never-takers* ($U_i = n$); and in case (4), compliance type U_i can only be *always-takers* ($U_i = a$) (Table 5.1). This is because of the monotonicity assumption, which excludes *defiers*.

So we define P^i_{ac}, P^i_{nc}, P^i_n and P^i_a as follows:

$$\begin{cases} P^i_{ac} & = Pr(U_i = c|(T_i(1), C_i(1))^{obs}, Z_i = 1, D_i(1) = 1; \boldsymbol{\theta}^k) \quad \text{for } R_i = 1 , \\ P^i_{nc} & = Pr(U_i = c|(T_i(0), C_i(0))^{obs}, Z_i = 0, D_i(0) = 0; \boldsymbol{\theta}^k) \quad \text{for } R_i = 1 , \\ P^i_n & = Pr(U_i = n|(T_i(1), C_i(1))^{obs}, Z_i = 1, D_i(1) = 0; \boldsymbol{\theta}^k) \quad \text{for } R_i = 1 , \\ P^i_a & = Pr(U_i = a|(T_i(0), C_i(0))^{obs}, Z_i = 0, D_i(0) = 1; \boldsymbol{\theta}^k) \quad \text{for } R_i = 1 . \end{cases} \tag{5.9.1}$$

Similarly, we use Q^i_{ac}, Q^i_{nc}, Q^i_n and Q^i_a to denote the related posterior probabilities for censored $R_i = 0$ associated with the aforementioned four cases, they are defined as the following:

$$\begin{cases} Q^i_{ac} & = Pr(U_i = c|(T_i(1), C_i(1))^{obs}, Z_i = 1, D_i(1) = 1; \boldsymbol{\theta}^k) \quad \text{for } R_i = 0 , \\ Q^i_{nc} & = Pr(U_i = c|(T_i(0), C_i(0))^{obs}, Z_i = 0, D_i(0) = 0; \boldsymbol{\theta}^k) \quad \text{for } R_i = 0, \\ Q^i_n & = Pr(U_i = n|(T_i(1), C_i(1))^{obs}, Z_i = 1, D_i(1) = 0; \boldsymbol{\theta}^k) \quad \text{for } R_i = 0, \\ Q^i_a & = Pr(U_i = a|(T_i(0), C_i(0))^{obs}, Z_i = 0, D_i(0) = 1; \boldsymbol{\theta}^k) \quad \text{for } R_i = 0 . \end{cases} \tag{5.9.2}$$

Note that when the event occurs (non-censored and denoted by $R_i = 1$), the observed data includes: failure time $T_i(z)$, censoring time ($C_i(z) > T_i(z)$), assigned treatment $Z_i = z$ and actual received treatment $D_i(z)$, and when an individual is censored (denoted by $R_i = 0$), the observed data includes: censoring time $C_i(z)$, censored failure time $T_i(z) > C_i(z)$, assigned treatment Z_i and actual received treatment $D_i(z)$.

The formulations for the posterior probabilities corresponding to the last

two cases, namely $Z_i = 1,\ D_i(1) = 0$; and $Z_i = 0,\ D_i(0) = 1$, are as follows:

$$\begin{cases} Pr(U_i = n \quad |(T_i(1),\, C_i(1))^{obs},\ Z_i = 1,\ D_i = 0;\ \boldsymbol{\theta}^k) = 1; & \text{for } R_i = 1; \\ Pr(U_i = a \quad |(T_i(0),\, C_i(0))^{obs},\ Z_i = 0,\ D_i = 1;\ \boldsymbol{\theta}^k) = 1; & \text{for } R_i = 1; \\ Pr(U_i = n \quad |(T_i(1),\, C_i(1))^{obs},\ Z_i = 1,\ D_i = 0;\ \boldsymbol{\theta}^k) = 1; & \text{for } R_i = 0; \\ Pr(U_i = a \quad |(T_i(0),\, C_i(0))^{obs},\ Z_i = 0,\ D_i = 1;\ \boldsymbol{\theta}^k) = 1; & \text{for } R_i = 0. \end{cases} \tag{5.9.3}$$

This is true, because no defiers exist and by the definition of compliance type U_i, we know that an individual is a *never-taker* if he/she is assigned to the treatment group $Z_i = 1$ but does not receive the treatment $D_i(1) = 0$ or is an *always-taker* if the individual is assigned to the control group $Z_i = 0$ but for some reason does receive the treatment.

Equation (5.9.3) gives the following equivalences:

$$\begin{cases} P_n^i = 1 & \text{for } \ i \in \mathcal{C}(1,0,1), \\ Q_n^i = 1 & \text{for } \ i \in \mathcal{C}(1,0,0), \\ P_a^i = 1 & \text{for } \ i \in \mathcal{C}(0,1,1), \\ Q_a^i = 1 & \text{for } \ i \in \mathcal{C}(0,1,0). \end{cases} \tag{5.9.4}$$

where $\mathcal{C}(z,d,r)$ denotes the subset of individuals, which is defined by the following: $\mathcal{C}(z,d,r) = \{i : Z_i = z,\ D_i = d,\ R_i = r\}$ for all $z,d,r \in \{1,0\}$. Now, we formulate the posterior probability of compliance type U_i for the first two cases, i.e. $Z_i = 1$ and $D_i(1) = 1$, and $Z_i = 0$ and $D_i(0) = 0$. Since both failure and censoring times are involved in the non-ignorable PPO-survival model, so too are their posterior probabilities. As per Bayes' theorem, these can be written as the formulae given in equations (5.9.5) and (5.9.6) (see next page). For the non-ignorable PPO-survival model, $P_{ac}^i,\ P_{nc}^i, Q_{ac}^i$ and Q_{nc}^i can then be written as the equivalences shown in equation (5.9.7) on page 108.

$$
\begin{aligned}
&Pr(U_i = u|(T_i(1),\ C_i(1))^{obs},\ Z_i = 1,\ D_i(1) = 1;\ \boldsymbol{\theta}^k) \\
&\qquad = \frac{f_u(1,t;\theta_{T,u1})G_u(1,t;\theta_{C,u1})\pi_{u,11}}{f_c(1,t;\theta_{T,c1})G_c(1,t;\theta_{C,c1})\pi_{c,11} + f_a(1,t;\theta_{T,a1})G_a(1,t;\theta_{C,a1})\pi_{a,11}}, \text{if } R_i = 1, T_i(1) = t; \\
&Pr(U_i = u|(T_i(1),\ C_i(1))^{obs}\ Z_i = 1,\ D_i(1) = 1;\ \boldsymbol{\theta}^k) \\
&\qquad = \frac{S_u(1,t;\theta_{T,u1})e_u(1,t;\theta_{C,u1})\pi_{u,11}}{S_c(1,t;\theta_{T,c1})e_c(1,t;\theta_{C,c1})\pi_{c,11} + S_a(1,t;\theta_{T,a1})e_a(1,t;\theta_{C,a1})\pi_{a,11}}, \text{if } R_i = 0, C_i(1) = t;
\end{aligned}
\tag{5.9.5}
$$

for $u \in \{c,\ a\}$; and

$$
\begin{aligned}
&Pr(U_i = u|(T_i(0),\ C_i(0))^{obs},\ Z_i = 0,\ D_i(0) = 0;\ \boldsymbol{\theta}^k) \\
&\qquad = \frac{f_u(0,t;\theta_{T,u0})G_u(1,t;\theta_{C,u0})\pi_{u,00}}{f_c(0,t;\theta_{C,c0})G_c(0,t;\theta_{C,c0})\pi_{c,00} + f_n(0,t;\theta_{T,n0})G_n(0,t;\theta_{C,n0})\pi_{n,00}},\ \text{if } R_i = 1,\ T_i(0) = t; \\
&Pr(U_i = u|(T_i(0),\ C_i(0))^{obs},\ Z_i = 0,\ D_i(0) = 0;\ \boldsymbol{\theta}^k) \\
&\qquad = \frac{S_u(0,t;\theta_{T,u0})e_u(0,t;\theta_{C,u0})\pi_{u,00}}{S_c(0,t;\theta_{T,c0})e_c(0,t;\theta_{C,c0})\pi_{c,00} + S_n(0,t;\theta_{T,n0})e_n(0,t;\theta_{C,n0})\pi_{n,00}},\ \text{if } R_i = 0, C_i(0) = t;
\end{aligned}
\tag{5.9.6}
$$

for $u \in \{c,\ n\}$.

$$
\begin{aligned}
P^i_{ac} &= \frac{f_c(1,t;\theta_{T,c1})G_c(1,t;\theta_{C,c1})\pi_{c,11}}{f_c(1,t;\theta_{T,c1})G_c(1,t;\theta_{C,c1})\pi_{c,11} + f_a(1,t;\theta_{T,a1})G_a(1,t;\theta_{C,a1})\pi_{a,11}}; \quad \text{for } i \in \mathcal{C}(1,1,1) \\
P^i_{nc} &= \frac{f_c(0,t;\theta_{T,c0})G_c(1,t;\theta_{C,c0})\pi_{c,00}}{f_c(0,t;\theta_{T,c0})G_c(0,t;\theta_{C,c0})\pi_{c,00} + f_{n0}(0,t;\theta_{T,n0})G_{n0}(0,t;\theta_{C,n0})\pi_{n,00}}; \quad \text{for } \ i \in \mathcal{C}(0,0,1) \\
Q^i_{ac} &= \frac{S_c(1,t;\theta_{T,c1})e_c(1,t;\theta_{C,c1})\pi_{c,11}}{S_c(1,t;\theta_{T,c1})e_c(1,t;\theta_{C,c1})\pi_{c,11} + S_a(1,t;\theta_{T,a1})e_a(1,t;\theta_{C,a1})\pi_{a,11}}; \quad \text{for } \ i \in \mathcal{C}(1,1,0) \\
Q^i_{nc} &= \frac{S_c(0,t;\theta_{T,c0})e_c(0,t;\theta_{C,c0})\pi_{c,00}}{S_c(0,t;\theta_{T,c0})e_c(0,t;\theta_{C,c0})\pi_{c,00} + S_n(0,t;\theta_{T,n0})e_n(0,t;\theta_{C,n0})\pi_{n,00}}; \quad \text{for } \ i \in \mathcal{C}(0,0,0).
\end{aligned}
\tag{5.9.7}
$$

For convenience, we use $W_i(u, \boldsymbol{\theta}^k)$ to denote the posterior probability of compliance type $Pr(U_i = u|(T_i, C_i)^{obs}, Z_i = z, D_i(z) = d, \boldsymbol{\theta}^k)$ given the observed data (and parameter values) for the non-ignorable PPO-survival model. That is:

$$W_i(u, \boldsymbol{\theta}^k) = Pr(U_i = u|(T_i, C_i)^{obs}, Z_i = z, D_i(z) = d; \boldsymbol{\theta}^k), \tag{5.9.8}$$

where $(T_i, C_i)^{obs}$ is same as given in equation (5.5.32), and indicates the observable part of survival outcomes $(T_i,\ C_i)$. $(T_i, C_i)^{obs}$ is equivalent to $T_i^{obs} = T_i(z)$ given $Z_i = z$, when $R_i = 1$; $(T_i, C_i)^{obs}$ is equivalent to $C_i^{obs} = C_i(z)$ given $Z_i = z$, when $R_i = 0$; and $\boldsymbol{\theta}^k$ represents the given value of the relevant model parameters, the components of which are:

$$(\rho_{T,c1},\ \rho_{T,c0},\ \rho_{T,n},\ \rho_{T,a},\ \alpha,\ \rho_{C,c1},\ \rho_{C,c0}, \rho_{C,n},\ \rho_{C,a},\ \alpha_C) \tag{5.9.9}$$

for the NIGN-WW model and

$$(\mu_{T,c1},\ \mu_{T,c0},\ \mu_{T,n},\ \mu_{T,a},\ \sigma,\ \mu_{C,c1},\ \mu_{C,c0}, \mu_{C,n},\ \mu_{C,a},\ \sigma_c) \tag{5.9.10}$$

for the NIGN-LL model.

The posterior probability $W_i(u, \theta^k)$ of the *ith* individual can then be written as follows:

$$W_i(u, \boldsymbol{\theta}^k) = \begin{cases} (P^i_{ac})^{I\{U_i=c\}}(1 - P^i_{ac});^{I\{U_i=a\}} & \text{for } i \in \mathcal{C}(1,1,1), \\ (P^i_{nc})^{I\{U_i=c\}}(1 - P^i_{nc});^{I\{U_i=n\}} & \text{for } i \in \mathcal{C}(0,0,1), \\ (Q^i_{ac})^{I\{U_i=c\}}(1 - Q^i_{ac});^{I\{U_i=a\}} & \text{for } \ i \in \mathcal{C}(1,1,0), \\ (Q^i_{nc})^{I\{U_i=c\}}(1 - Q^i_{nc});^{I\{U_i=n\}} & \text{for } \ i \in \mathcal{C}(0,0,0), \\ P^i_n = 1; & \text{for } \ i \in \mathcal{C}(1,0,1), \\ Q^i_n = 1; & \text{for } \ i \in \mathcal{C}(1,0,0), \\ P^i_a = 1; & \text{for } \ i \in \mathcal{C}(0,1,1), \\ Q^i_a = 1; & \text{for } \ i \in \mathcal{C}(0,1,0); \end{cases} \tag{5.9.11}$$

where $I\{\}$ is an indicator function, and $P^i_{ac},\ P^i_{nc}, Q^i_{ac}$ and Q^i_{nc} denote the posterior probabilities for the specific cases, given by equation (5.9.7) for the

non-ignorable PPO-survival model. Recall that $\mathcal{C}(z,d,r)$ denotes the subset of individuals $\mathcal{C}(z,d,r) = \{i:\ Z_i = z,\ D_i = d,\ R_i = r\}$ for all $z,d,r \in \{1,0\}$.

5.9.2 Structures of the expectation function $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^k)$

After defining the posterior probability, $W_i(u, \boldsymbol{\theta}^k)$, which is given in equation (5.9.11), the expectation equation, Q-function $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^k)$ specified in equation (5.5.33) can then be re-written as:

$$\begin{aligned} Q^i(\boldsymbol{\theta},\boldsymbol{\theta}^k) &= \sum_{u\in\{c,n,a\}} W_i(u,\boldsymbol{\theta}^k)\{log(l^F(\boldsymbol{\theta})\} \\ &= \sum_{u\in\{c,n,a\}} W_i(u,\boldsymbol{\theta}^k)\{log(l(\boldsymbol{\theta},\, u)\}, \end{aligned} \tag{5.9.12}$$

where $l^F(\boldsymbol{\theta})$ is given in equation (5.5.15).

After substituting equations (5.9.7), (5.9.11) and (5.5.15) into equation (5.9.12), $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^k)$ can be specified as in equation (5.9.13) overleaf.

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^k) = \sum_i Q^i(\boldsymbol{\theta}, \boldsymbol{\theta}^k) =$$

$$\begin{cases} \sum_{i \in \mathcal{C}(1,1,1)} \left\{ log\left(f_c^i(1,t;\theta_{T,c1}) G_c^i(1,t;\theta_{C,c1}) \pi_{c,11}\right) P_{ac}^i + log\left(f_a^i(1,t;\theta_{T,a}) G_a^i(1,t;\theta_{C,a}) \pi_{a,11}\right) (1 - P_{ac}^i) \right\} \\ \sum_{i \in \mathcal{C}(1,1,0)} \left\{ log\left(S_c^i(1,t;\theta_{T,c1}) e_c^i(1,t;\theta_{C,c1}) \pi_{c,11}\right) Q_{ac}^i + log\left(S_a^i(1,t;\theta_{T,a}) e_a^i(1,t;\theta_{C,a}) \pi_{a,11}\right) (1 - Q_{ac}^i) \right\} \\ \sum_{i \in \mathcal{C}(0,0,1)} \left\{ log\left(f_c^i(0,t;\theta_{T,c0}) G_c^i(0,t;\theta_{C,c0}) \pi_{c,00}\right) P_{nc}^i + log\left(f_n^i(0,t;\theta_{T,n}) G_n^i(0,t;\theta_{C,n}) \pi_{n,00}\right) (1 - P_{nc}^i) \right\} \\ \sum_{i \in \mathcal{C}(0,0,0)} \left\{ log\left(S_c^i(0,t;\theta_{T,c0}) e_c^i(0,t;\theta_{C,c0}) \pi_{c,00}\right) Q_{nc}^i + log\left(S_n^i(0,t;\theta_{T,n}) e_n^i(0,t;\theta_{C,n}) \pi_{n,00}\right) (1 - Q_{nc}^i) \right\} \\ \sum_{i \in \mathcal{C}(1,0,1)} log\left(f_n^i(1,t;\theta_{T,n}) G_n^i(1,t;\theta_{C,n}) \pi_{n,10}\right) \sum_{i \in \mathcal{C}(1,0,0)} log\left(S_n^i(1,t;\theta_{T,n}) e_n^i(1,t;\theta_{C,n}) \pi_{n,10}\right) \\ \sum_{i \in \mathcal{C}(0,1,1)} log\left(f_a^i(0,t;\theta_{T,a}) G_a^i(0,t;\theta_{C,a}) \pi_{a,01}\right) \sum_{i \in \mathcal{C}(0,1,0)} log\left(S_a^i(0,t;\theta_{T,a}) e_a^i(0,t;\theta_{C,a}) \pi_{a,01}\right). \end{cases} \tag{5.9.13}$$

5.9.3 Maximization of the expectation function $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^k)$

As stated in Section 5.5.5, the M-step aims to maximise the expectation function given in equation (5.9.13). To realise this goal, the score equation (McLachlan & Krishnan, 1997; Casella & Berger, 2002) is often used in the M-step. Specifically, for the *ith* individual, we take the first partial derivative of $Q^i(\theta, \theta^k)$, with respect to the parameter θ in equation (5.9.13) for the non-ignorable PPO model. Then the i^{th} score equation is:

$$\mathfrak{S}_i^{obs}(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} Q^i(\boldsymbol{\theta}, \boldsymbol{\theta}^k) = \sum_{u \in \{c,n,a\}} W_i(u, \boldsymbol{\theta}^k) \frac{\partial}{\partial \boldsymbol{\theta}} logl_i(\boldsymbol{\theta}, u). \tag{5.9.14}$$

As before, let $S_{i,\boldsymbol{\theta}}(u, \boldsymbol{\theta})$ denote $\frac{\partial}{\partial \boldsymbol{\theta}} logl_i(\theta, u)$. Then the score equation for the i^{th} individual can be expressed as:

$$\mathfrak{S}_i^{obs}(\boldsymbol{\theta}) = \sum_u W_i(u, \theta^k) S_{i,\boldsymbol{\theta}}(u, \boldsymbol{\theta}). \tag{5.9.15}$$

All parameters can then be estimated simultaneously by solving:

$$\sum_i^n \mathfrak{S}_i^{obs}(\boldsymbol{\theta}) = 0. \tag{5.9.16}$$

Let $\boldsymbol{\theta}^{(k+1)}$ denote the solution of equation (5.9.16). We then substitute $\boldsymbol{\theta}^{(k)}$ in $Q(\boldsymbol{\theta},\ \boldsymbol{\theta}^{(k)})$ with $\boldsymbol{\theta}^{(k+1)}$ and repeat the M-step until convergence.

The models proposed in this thesis use a built-in function in MATLAB (7.4.0, R2007a) (Higham & Higham, 2000). This function, **fminsearch**, is based on the Nelder-Mead method (Lagarias et al., 1998) for multidimensional unconstrained minimisation. We use it to minimise the negative expectation function $-Q(\boldsymbol{\theta},\ \boldsymbol{\theta}^{(k)})$, which is equivalent to maximising the expectation function $Q(\boldsymbol{\theta},\ \boldsymbol{\theta}^{(k)})$.

5.9.4 Estimation of the standard errors

The standard error for the estimated model parameter

The EM algorithm relies only on the convergence of iterative procedures, in which no gradient matrices are derived. As such, the EM algorithm does

not provide the information matrix required for traditional calculation of the asymptotic variance-covariance matrix of the MLEs, from either the expected information matrix or the observed information matrix (OIM). Consequently, no standard errors of the MLEs can be obtained, even after possible adaptation of the EM algorithm.

Louis (1982), however, proposed an approach for calculating the observed information matrix (OIM) when the EM algorithm is used to find MLEs for so-called incomplete data problems. Louis's approach (Louis, 1982) only requires computation of the complete-data gradient vector or computation of the second derivative matrix, (which is not associated with the incomplete-data likelihood function). After obtaining the OIM, the covariance matrix can then be calculated easily. Only the inverse of the OIM is then required. The standard error of the MLEs vectors can thus be obtained easily, in that the standard error is the square root of the diagonals of the covariance matrix, where the latter is the inverse of the observed information matrix (OIM).

Louis's method[1]

Let $Data_{full}$ and $Data_{obs}$ denote the complete (full) and incomplete (observed) data, respectively. The former includes the observed time $(T,\ C)^{obs}$, (which is equivalent to $(Y(z), R(z))$ in survival analysis, where $Y(z) = min(T(z), C(z))$ and $R(z)$ is the censoring indicator), assigned treatment $Z = z$, received treatment $D(z)$ and the compliance type U. Thereby $Data_{full} = ((T,\ C)^{obs},\ Z = z,\ D(z),\ U)$. As compliance type U is unobservable, we have $Data_{obs} = ((T,\ C)^{obs},\ Z = z,\ D(z))$. By following equation (3.3) in Louis (1982), the observed information matrix (OIM), $I_{obs} = I(\hat{\boldsymbol{\theta}})$, which can be viewed as the matrix at the MLEs, where matrix I can be expressed in two parts: the conditional expected full data OIM, I_{full}, and the expected information for the conditional distribution of the complete data $I_{full|obs}$. This is represented as follows:

$$I_{obs} = I(\hat{\boldsymbol{\theta}}) = I_{full}(\hat{\boldsymbol{\theta}}) - I_{full|obs}(\hat{\boldsymbol{\theta}}). \tag{5.9.17}$$

[1]Equation (3.2′) of Louis (1982) is not strictly correct. We believe that the rule of non-exchangeable vector order in the product was ignored. Details of our correction are given in Appendix A.

Under the independent and identically distributed (i.i.d) assumption, these two matrices I_{full} and $I_{full|obs}$ can be specified as:

$$I_{full} = E_u\{-\sum_{i=1}^{n} \frac{\partial^2}{\partial \boldsymbol{\theta}^2} logL_i(\boldsymbol{\theta}, u)|(T_i,\ C_i)^{obs},\ Z_i = z,\ D_i(z) = d,\ U_i = u\} \tag{5.9.18}$$

$$\begin{aligned} & I_{full|obs} \\ &= \sum_i E_u \left\{ \left(\frac{\partial}{\partial \boldsymbol{\theta}} logl_i^F(\boldsymbol{\theta}) \right)^T \left(\frac{\partial}{\partial \boldsymbol{\theta}} logl_i^F(\boldsymbol{\theta}) \right) |(T_i,\ C_i)^{obs},\ Z_i = z,\ D_i(z) = d,\ U_i = u \right\} \\ &\quad + \sum_{i \neq j} \left\{ \begin{array}{l} E_u \left(\frac{\partial}{\partial \boldsymbol{\theta}} logl_i^F(\boldsymbol{\theta}) |(T_i,\ C_i)^{obs},\ Z_i = z,\ D_i(z) = d,\ U_i = u \right)^T \times \\ E_u \left(\frac{\partial}{\partial \boldsymbol{\theta}} logl_j^F(\boldsymbol{\theta}) |(T_i,\ C_i)^{obs},\ Z_i = z,\ D_i(z) = d,\ U_i = u \right) \end{array} \right\} \end{aligned} \tag{5.9.19}$$

where ^{T} denotes transpose. Consequently, the observed information matrix I_{obs} can be expressed[2], after correcting the classical method of Louis (1982) as:

$$\begin{aligned} I_{obs} &= I(\hat{\boldsymbol{\theta}}) \\ &= \sum_{i=1}^{n} E_u \left\{ -\frac{\partial^2}{\partial \boldsymbol{\theta}^2} \left(logl_i(\boldsymbol{\theta}, u) |(T_i,\ C_i)^{obs},\ Z_i = z,\ D_i(z) = d \right) \right\}_{|\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}} \\ &\quad - \sum_i E_u \left\{ \left(\frac{\partial}{\partial \boldsymbol{\theta}} logl_i(\boldsymbol{\theta}, u) \right)^T \left(\frac{\partial}{\partial \boldsymbol{\theta}} logl_i(\theta, u) \right) |(T_i,\ C_i)^{obs},\ Z_i = z,\ D_i(z) = d \right\}_{|\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}} \\ &\quad - \sum_{i \neq j} \left\{ \begin{array}{l} E_u \left(\frac{\partial}{\partial \theta} logl_i(\theta, u) |(T_i,\ C_i)^{obs},\ Z_i = z,\ D_i(z) = d \right)^T \times \\ E_u \left(\frac{\partial}{\partial \theta} logl_j(\theta, u) |(T_j,\ C_j)^{obs},\ Z_j = z,\ D_j(z) = d \right) \end{array} \right\}_{|\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}} . \end{aligned} \tag{5.9.20}$$

Equation (5.9.20) (also see equation (A.4) in Appendix A) guarantees that the OIM is symmetric. (Details are given on page 115.) As the inverse of I_{obs} is the asymptotic variance-covariance matrix of the MLEs, $\hat{\boldsymbol{\theta}}$, this can be denoted as $\mathcal{V}(\hat{\boldsymbol{\theta}}) = I^{-1}(\hat{\boldsymbol{\theta}})$. Its diagonal vector is the variance vector associated with each of the components of $\hat{\boldsymbol{\beta}}$ and is denoted by $SE_{\hat{\boldsymbol{\theta}}}$. The standard error of $\hat{\boldsymbol{\theta}}$ can then be calculated directly by taking the square root of $\sigma^2_{\hat{\theta}}$, where $\sigma^2_{\hat{\boldsymbol{\theta}}} = \{\mathcal{V}(\hat{\boldsymbol{\theta}})\}_{ii}\, (i = 1, 2, \cdots, p)$, for p parameters.

---

[2]This equation can be regarded as our correction of equation (3.2') of Louis (1982). Louis's subscripts should specify that $i \neq j$ rather than $i < j$.

The observed information matrix (OIM)

Note that the first term in equation (5.9.17) is equivalent to the second order partial derivative of Q-function $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^k)$, which can be written as $\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^2}$ and $\boldsymbol{\theta}$ is a interest parameter vector with length p. If we write $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)$, the element at the *ith* row and the *jth* column in the matrix $\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^2}$, the second order derivative matrix of Q-function $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$ with respect to the *ith* and *jth* components of the interest parameter $\boldsymbol{\theta}$, θ_i and θ_j , can be written as follows:

$$\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^2} = \left[\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})}{\partial \theta_i \partial \theta_j}\right]_{ij} \quad \text{for } i \text{ and } j \text{ from 1 to } p. \tag{5.9.21}$$

Since in the PPO-survival model, the interest parameter vector $\boldsymbol{\theta}$, given in equation (5.5.31), consists of the parameters associated with the distributions of failure and censoring times for three subgroups. This leads to the fact of that majority of elements in the matrix $\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^2}$ are zeros. For instance, the second order partial derivative of the Q-function,$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})$, $\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)})}{\partial \boldsymbol{\theta}^2}$ with respect to two scale parameters $\theta_1 = \rho_{T,c1}$ and $\theta_2 = \rho_{T,c0}$ (in Weibull PPO-survival model) is zero. That is:

$$\frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})}{\partial \theta_i \partial \theta_j} = \frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})}{\partial \theta_1 \partial \theta_2} = 0 \tag{5.9.22}$$

As no an explicit relationship exists between θ_1 and θ_2, which represent the scale parameter from two different subgroups here.

Even though in the same subgroup, however, with respect to the parameters associated with the distributions of failure and censoring times, $\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)})}{\partial \theta_i \partial \theta_j}$ can also be zero. For example, let $i = 1$ and $j = 5$, in Weibull PPO-survival model we have $\theta_1 = \rho_{T,c1}$ and $\theta_5 = \rho_{C,c1}$, where $\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)})}{\partial \theta_1 \partial \theta_5} = 0$ always holds. This is because there is not an explicit relationship between these two parameters $\rho_{T,c1}$ and $\rho_{C,c1}$ which are from two different distributions. Recall that $\rho_{T,c1}$ denotes the scale parameter of Weibull for failure time in and $\rho_{C,c1}$ denotes the scale parameter of Weibull for censoring time in subgroup1).

Therefore, in the PPO-survival model (including both Weibull and Log-normal form), the first term of equation (5.9.20) can be simplified directly without knowing the estimated value of parameter $\hat{\boldsymbol{\theta}}$. The simplified form of the matrix is given in Table 5.5 (page 116) where θ_i and θ_j denote any components of parameter $\boldsymbol{\theta}$, for the specific model, say the NIGN-WW or the NIGN-LL models, $\boldsymbol{\theta}$ is de-

fined in Table 5.4 (on page 104). Note that Table 5.5 shows that in the matrix, $[\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)})}{\partial\theta_i\partial\theta_j}]_{10\times10}$, only few of elements are not zero (denoted by $*$ in the table) which can be calculated directly through the standard likelihood function. In this thesis, the standard likelihood function were given as either Weibull or log-normal form.

Table 5.5: The patten of second order partial derivative matrix of $Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)})$.

$\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)})}{\partial\theta_i\partial\theta_j}$	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}
θ_1	$*$	0	0	0	$*$	0	0	0	0	0
θ_2	0	$*$	0	0	$*$	0	0	0	0	0
θ_3	0	0	$*$	0	$*$	0	0	0	0	0
θ_4	0	0	0	$*$	$*$	0	0	0	0	0
θ_5	$*$	$*$	$*$	$*$	$*$	0	0	0	0	0
θ_6	0	0	0	0	0	$*$	0	0	0	$*$
θ_7	0	0	0	0	0	0	$*$	0	0	$*$
θ_8	0	0	0	0	0	0	0	$*$	0	$*$
θ_9	0	0	0	0	0	0	0	0	$*$	$*$
θ_{10}	0	0	0	0	0	$*$	$*$	$*$	$*$	$*$

All these components of $\boldsymbol{\theta}$ are defined in Table 5.4 on page 104.

From Table 5.5, it is easy to see that $[\frac{\partial^2 Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)})}{\partial\theta_i\partial\theta_j}]_{10\times10}$, the second order partial derivative matrix of Q-function $Q(\boldsymbol{\theta},\boldsymbol{\theta}^{(k)})$ is a symmetric matrix. In fact, the second term in equation (5.9.20) also denotes a symmetric matrix as it is a summation of a product of a vector and it's transpose; the third term in the same equation represents a symmetric too, even though two different vectors were involved in the product term. This is because that the later summation is taken over on $i \neq j$ only, which means that those single matrix that satisfied $i = j$ was not included in the summation. Therefore, the OIM, defined by equation (5.9.20) is really a symmetric matrix as their three components are all symmetric matrices.

The standard error (SE) of $\widehat{CACE}(t)$ *and* $\widehat{ACE}(t)$

As noted in Chapter 3, estimating $CACE(t)$ and $ACE(t)$ rather than estimating the model parameters $\boldsymbol{\theta}_{T,uz}$ may be significantly more valuable and more interpretable. The standard errors corresponding to the estimated $\widehat{CACE}(t)$ and $\widehat{ACE}(t)$ have not been derived to date, to the best of the author's knowledge.

We now consider the derivation of formulae for calculating the standard errors of the estimated $CACE(t)$ and $ACE(t)$ via the estimated model parameters $\boldsymbol{\theta}_{T,uz}$. Recall that $CACE(t)$ was defined as the difference in survival values between compliers in the treated group and the control group (Section 3.3). By following the notation of the customary survival function (for subgroups), $CACE(t)$ can thus be re-written as follows:

$$CACE(t) = S_{c1}(t; \boldsymbol{\theta}_{T,c1}) - S_{c0}(t; \boldsymbol{\theta}_{T,c0}). \tag{5.9.23}$$

According to the property of variance of linear combinations (Theorem 4.5.6, p.171, Casella & Berger (2002)), the variance of $CACE(t)$ can be easily expressed as:

$$\begin{aligned} Var(CACE(t)) = Var(S_{c1}(t; \boldsymbol{\theta}_{T,c1})) + Var(S_{c0}(t; \boldsymbol{\theta}_{T,c0})) \\ - 2Cov(S_{c1}(t; \boldsymbol{\theta}_{T,c1}), S_{c0}(t; \boldsymbol{\theta}_{T,c0})), \end{aligned} \tag{5.9.24}$$

where $S_{cz}(t; \boldsymbol{\theta}_{T,cz})$ (for $z = 1$ or 0) denotes the survival function at time t associated with the subgroup of individuals who have compliance type $U_i = c$ and who have been assigned into the group of $Z_i = z$ ($z = 1$ for the treated group and $z = 0$ for the control group).

Note that except for the exponential distribution, most common parametric models for survival outcomes, such as the Weibull, Gamma, log-normal, and the log-logistic models have two parameters (Klein, 1997; Ibrahim et al., 2001). Let $\boldsymbol{\theta}_{T,uz} = (\theta^{(1)}_{T,uz}, \theta^{(2)}_{T,uz})$, and we treat $Var(S_{uz}(t; \boldsymbol{\theta}_{T,uz}))$ as the variance of a function of two variables $(\theta^{(1)}_{T,uz}, \theta^{(2)}_{T,uz})$. Then the $Cov(S_{c1}(t; \boldsymbol{\theta}_{T,c1}), S_{c0}(t; \boldsymbol{\theta}_{T,c0}))$ has to be viewed as the covariance function of two variables $(\theta^{(1)}_{T,uz}, \theta^{(2)}_{T,uz})$. As such, we deal with the variance and covariance of functions of multiple random variables. The δ method (Klein, 1997) (see below) can then be used to derive formulae for such approximations.

The δ method

The δ method, also known as the method of statistical differential (Klein, 1997), is based on a Taylor series expansion of a continuous function of the MLEs of a vector of parameters $(\theta^{(1)}_{T,uz}, \theta^{(2)}_{T,uz})$ for $u = c$ and $v = 1$ or 0. For convenience, let $f(x, y)$ denote a function of two variables, e.g. x and y; and suppose that at the point (x_0, y_0), the first partial derivative of $f(x, y)$ with respect to x and y exist. These are denoted by $f'_x(x_0, y_0)$ and $f'_y(x_0, y_0)$. Then the first order Taylor series

expansion at the point (x_0, y_0) has the form as follows:

$$f(x,y) = f(x_0,y_0) + f'_x(x_0,y_0)(x-x_0) + f'_y(x_0,y_0)(y-y_0). \tag{5.9.25}$$

If x and y are both random variables with mean values, $\bar{x}$ and $\bar{y}$, then setting $x_0 = \bar{x}$ and $y_0 = \bar{y}$, allows equation (5.9.25) to be expressed as:

$$f(x,y) = f(\bar{x},\bar{y}) + f'_x(\bar{x},\bar{y})(x-\bar{x}) + f'_y(\bar{x},\bar{y})(y-\bar{y}). \tag{5.9.26}$$

Variance of the function of two variables

From equation (5.9.26), it is straightforward to observe that the function of two random variables, $f(x,y)$, has the mean value $f(\bar{x},\bar{y})$, as the last two terms in equation (5.9.26) are zeros. By following the properties of variance (Theorem 4.5.6, p.171 Casella & Berger (2002)), we have:

$$\begin{aligned} Var(f(x,y)) = {} & f'_x(\bar{x},\bar{y})^2 Var(x) + f'_y(\bar{x},\bar{y})^2 Var(y) \\ & + 2f'_x(\bar{x},\bar{y}) f'_y(\bar{x},\bar{y}) Cov(x,y), \end{aligned} \tag{5.9.27}$$

because variances for constants are zeros, we have $Var(f(\bar{x},\bar{y})) = 0$, $Var(\bar{x}) = 0$ and $Var(\bar{y}) = 0$.

By following equation (5.9.27), we can write:

$$\begin{aligned} & Var(S_{uz}(t;\boldsymbol{\theta}_{T,uz})) = Var(S_{uz}(t;\theta^{(1)}_{T,uz},\theta^{(2)}_{T,uz})) \\ & = \left[\frac{\partial S_{uz}(t;\theta^{(1)}_{T,uz},\theta^{(2)}_{T,uz})}{\partial \theta^{(1)}_{T,uz}}\right]^2_{(\widehat{\theta}^{(1)}_{T,uz},\widehat{\theta}^{(2)}_{T,uz})} var(\theta^{(1)}_{T,uz}) \\ & \quad + \left[\frac{\partial S_{uz}(t;\theta^{(1)}_{T,uz},\theta^{(2)}_{T,uz})}{\partial \theta^{(2)}_{T,uz}}\right]^2_{(\widehat{\theta}^{(1)}_{T,uz},\widehat{\theta}^{(2)}_{T,uz})} var(\theta^{(2)}_{T,uz}) \\ & \quad + \left[\frac{\partial S_{uz}(t;\theta^{(1)}_{T,uz},\theta^{(2)}_{T,uz})}{\partial \theta^{(1)}_{T,uz}}\right]_{(\widehat{\theta}^{(1)}_{T,uz},\widehat{\theta}^{(2)}_{T,uz})} \left[\frac{\partial S_{uz}(t;\theta^{(1)}_{T,uz},\theta^{(2)}_{T,uz})}{\partial \theta^{(2)}_{T,uz}}\right]_{(\widehat{\theta}^{(1)}_{T,uz},\widehat{\theta}^{(2)}_{T,uz})} cov(\theta^{(1)}_{T,uz},\theta^{(2)}_{T,uz}). \end{aligned} \tag{5.9.28}$$

The approximate formula of $Var(S_u(v,t;\theta^{(1)}_{T,uz},\theta^{(2)}_{T,uz}))$, the variance of the survival function, can then be expressed by equation (5.9.28).

Covariance of functions of two variables

Similarly, if $h(u,v)$ denotes another function of two random variables u and v, with mean values $\bar{u}$ and $\bar{v}$, then the function $h(u,v)$ has a mean value $h(\bar{u},\bar{v})$.

By the definition of covariance, the covariance between $f(x,y)$ and $h(u,v)$ is then the expectation of the product of two factors $(f(x,y)-f(\bar{x},\bar{y}))$ and $(h(u,v)-h(\bar{u},\bar{v}))$; that is:

$$Cov(f(x,y),h(u,v)) = E\{(f(x,y)-f(\bar{x},\bar{y}))(h(u,v)-h(\bar{u},\bar{v}))\}. \tag{5.9.29}$$

By first order Taylor series expansion, equation (5.9.29) can be expressed by first order partial derivative values at the mean value points, i.e.:

$$\begin{aligned} &Cov(f(x,y),h(u,v)) \\ &\quad = E\{f'_x(\bar{x},\bar{y})(x-\bar{x}) + f'_y(\bar{x},\bar{y})(y-\bar{y}))(h'_u(\bar{u},\bar{v})(u-\bar{u}) + h'_v(\bar{u},\bar{v})(v-\bar{v}))\}. \end{aligned} \tag{5.9.30}$$

Furthermore, the covariance between $f(x,y)$ and $h(u,v)$ can be simplified to

$$\begin{aligned} Cov(f(x,y),h(u,v)) &= f'_x(\bar{x},\bar{y})h'_u(\bar{u},\bar{v})Cov(x,u) + f'_y(\bar{x},\bar{y})h'_u(\bar{u},\bar{v})Cov(y,u) \\ &\quad + f'_x(\bar{x},\bar{y})h'_v(\bar{u},\bar{v})Cov(x,v) + f'_y(\bar{x},\bar{y})h'_v(\bar{u},\bar{v})Cov(y,v). \end{aligned} \tag{5.9.31}$$

By following equation (5.9.31), the covariance between survival functions, say $S_{c1}(t;\theta_{T,c1})$ and $S_{c0}(t;\theta_{T,c0})$, can then be expressed as follows:

$$\begin{aligned} &Cov(S_{c1}(t;\boldsymbol{\theta}_{T,c1}), S_{c0}(t;\boldsymbol{\theta}_{T,c0})) = \\ &\quad \left[\frac{\partial S_{c1}(t;\theta^{(1)}_{T,c1},\theta^{(2)}_{T,c1})}{\partial \theta^{(1)}_{T,c1}}\right]_{(\widehat{\theta}^{(1)}_{T,c1},\widehat{\theta}^{(2)}_{T,c1})} \left[\frac{\partial S_{c0}(t;\theta^{(1)}_{T,c0},\theta^{(2)}_{T,c0})}{\partial \theta^{(1)}_{T,c0}}\right]_{(\widehat{\theta}^{(1)}_{T,c0},\widehat{\theta}^{(2)}_{T,c0})} cov(\theta^{(1)}_{T,c1},\theta^{(1)}_{T,c0}) \\ &\quad + \left[\frac{\partial S_{c1}(t;\theta^{(1)}_{T,c1},\theta^{(2)}_{T,c1})}{\partial \theta^{(2)}_{T,c1}}\right]_{(\widehat{\theta}^{(1)}_{T,c1},\widehat{\theta}^{(2)}_{T,c1})} \left[\frac{\partial S_{c0}(t;\theta^{(1)}_{T,c0},\theta^{(2)}_{T,c0})}{\partial \theta^{(1)}_{T,c0}}\right]_{(\widehat{\theta}^{(1)}_{T,c0},\widehat{\theta}^{(2)}_{T,c0})} cov(\theta^{(2)}_{T,c1},\theta^{(1)}_{T,c0}) \\ &\quad + \left[\frac{\partial S_{c1}(t;\theta^{(1)}_{T,c1},\theta^{(2)}_{T,c1})}{\partial \theta^{(1)}_{T,c1}}\right]_{(\widehat{\theta}^{(1)}_{T,c1},\widehat{\theta}^{(2)}_{T,c1})} \left[\frac{\partial S_{c0}(t;\theta^{(1)}_{T,c0},\theta^{(2)}_{T,c0})}{\partial \theta^{(2)}_{T,c0}}\right]_{(\widehat{\theta}^{(1)}_{T,c0},\widehat{\theta}^{(2)}_{T,c0})} cov(\theta^{(1)}_{T,c1},\theta^{(2)}_{T,c0}) \\ &\quad + \left[\frac{\partial S_{c1}(t;\theta^{(1)}_{T,c1},\theta^{(2)}_{T,c1})}{\partial \theta^{(2)}_{T,c1}}\right]_{(\widehat{\theta}^{(1)}_{T,c1},\widehat{\theta}^{(2)}_{T,c1})} \left[\frac{\partial S_{c0}(t;\theta^{(1)}_{T,c0},\theta^{(2)}_{T,c0})}{\partial \theta^{(2)}_{T,c0}}\right]_{(\widehat{\theta}^{(1)}_{T,c0},\widehat{\theta}^{(2)}_{T,c0})} cov(\theta^{(2)}_{T,c1},\theta^{(2)}_{T,c0}). \end{aligned} \tag{5.9.32}$$

Equations (5.9.28) and (5.9.32) both can be simplified via vector operation as follows:

$$\begin{aligned} &Var\left(S_{cz}(t;\boldsymbol{\theta}_{T,cz})\right) \\ &= \left[\frac{\partial S_{cz}(t;\boldsymbol{\theta}_{T,cz})}{\partial \theta^{(1)}_{T,cz}}, \frac{\partial S_{cz}(t;\boldsymbol{\theta}_{T,cz})}{\partial \theta^{(2)}_{T,cz}}\right] \Sigma_1 \left[\frac{\partial S_{cz}(t;\theta_{T,cz})}{\partial \theta^{(1)}_{T,cz}}, \frac{\partial S_{cz}(t;\theta_{T,cz})}{\partial \theta^{(2)}_{T,cz}}\right]^T, \end{aligned} \tag{5.9.33}$$

and

$$\begin{aligned} &Cov\left(S_{c1}(t;\theta_{T,c1}), S_{c0}(t;\theta_{T,c0})\right) \\ &= \left[\frac{\partial S_{c1}(t;\theta_{T,c1})}{\partial \theta^{(1)}_{T,c1}}, \frac{\partial S_{c1}(t;\theta_{T,c1})}{\partial \theta^{(2)}_{T,c1}}\right] \Sigma_2 \left[\frac{\partial S_{c0}(t;\theta_{T,c0})}{\partial \theta^{(1)}_{T,c0}}, \frac{\partial S_{c0}(t;\theta_{T,c0})}{\partial \theta^{(2)}_{T,c0}}\right];^T \end{aligned} \tag{5.9.34}$$

where Σ_1 and Σ_2 represent the covariance matrix of all relevant parameters. More specifically, Σ_1 and Σ_2 can be written as:

$$\Sigma_1 = \begin{pmatrix} var(\theta^{(1)}_{T,cz}) & cov(\theta^{(1)}_{T,cz}, \theta^{(2)}_{T,cz}) \\ cov(\theta^{(2)}_{T,cz}, \theta^{(1)}_{T,cz}) & var(\theta^{(2)}_{T,cz}) \end{pmatrix} \tag{5.9.35}$$

and

$$\Sigma_2 = \begin{pmatrix} cov(\theta^{(1)}_{T,c1}, \theta^{(1)}_{T,c0}) & cov(\theta^{(1)}_{T,c1}, \theta^{(2)}_{T,c0}) \\ cov(\theta^{(2)}_{T,c1}, \theta^{(1)}_{T,c0}) & cov(\theta^{(2)}_{T,c1}, \theta^{(2)}_{T,c0}) \end{pmatrix}. \tag{5.9.36}$$

Then the standard error of $CACE(t)$ can be calculated by taking the square root of the variance:

$$SE_{CACE(t)} = \sqrt{Var(CACE(t))}. \tag{5.9.37}$$

Since we have $ACE(t) = \pi_c CACE(t)$, the variance of $ACE(t)$ can then be calculated by the following equation:

$$\begin{aligned} Var(ACE(t)) &= Var(\pi_c CACE(t)) \\ &= Var\left(\pi_c S_{c1}(t;\theta_{T,c1}) - \pi_c S_{c0}(t;\theta_{T,c0})\right) \\ &= Var\left(\pi_c S_{c1}(t;\theta_{T,c1})\right) + Var\left(\pi_c S_{c0}(t;\theta_{T,c0})\right) \\ &\quad - 2cov(\pi_c S_{c1}(t;\theta_{T,c1}), \pi_c S_{c0}(t;\theta_{T,c0})). \end{aligned} \tag{5.9.38}$$

Note that as $\theta_{T,c1} = (\theta^1_{T,c1}, \theta^2_{T,c1})$ and $\theta_{T,c0} = (\theta^1_{T,c0}, \theta^2_{T,c0})$, the two terms in the last line of equation (5.9.38), namely $\pi_c S_{c1}(t;\theta_{T,c1})$ and $\pi_c S_{c0}(t;\theta_{T,c0})$, can both be

viewed as functions of three variables, which are (π_c, $\theta^1_{T,c1}$, $\theta^2_{T,c1}$) (in $\pi_c S_{c1}(t;\theta_{T,c1})$) and ($\pi_c$, $\theta^1_{T,c0}$, $\theta^2_{T,c0}$) in $\pi_c S_{c0}(t;\theta_{T,c0})$.

By following the logic used to derive the formulations of the variance (equation (5.9.33)) and covariance (equation (5.9.34)) of the survival function, ($S_{uz}(t;\theta_{T,uz})$ (for $z = 1$ and 0)), the variance of $\pi_c S_{uz}(t;\theta_{T,uz})$, denoted by $Var(\pi_c S_{uz}(t;\theta_{T,uz}))$, can then be expressed as follows:

$$\begin{aligned} &Var\left(\pi_c S_{cz}(t;\boldsymbol{\theta}_{T,uz})\right) \\ &= \left\{ \begin{array}{l} \left[S_{cz}(t;\theta_{T,cz}),\ \pi_c \frac{\partial S_{cz}(t;\theta_{T,cz})}{\partial \theta^{(1)}_{T,cz}},\ \pi_c \frac{\partial S_{cz}(t;\theta_{T,cz})}{\partial \theta^{(2)}_{T,cz}}\right] \Sigma_3 \\ \left[S_{cz}(t;\theta_{T,cz}),\ \pi_c \frac{\partial S_{cz}(t|\theta_{T,cz})}{\partial \theta^{(1)}_{T,cz}},\ \pi_c \frac{\partial S_{cz}(t;\theta_{T,cz})}{\partial \theta^{(2)}_{T,cz}}\right]^T, \end{array} \right\} \end{aligned} \tag{5.9.39}$$

where Σ_3 is given by:

$$\Sigma_3 = \begin{pmatrix} var(\theta^{(1)}_{T,cz}) & cov(\theta^{(1)}_{T,cz}, \theta^{(2)}_{T,cz}) & cov(\theta^{(1)}_{T,cz}, \pi_c) \\ cov(\theta^{(2)}_{T,cz}, \theta^{(1)}_{T,cz}) & var(\theta^{(2)}_{T,cz}) & cov(\theta^{(2)}_{T,cz}, \pi_c) \\ cov(\pi_c, \theta^{(1)}_{T,cz}) & cov(\pi_c, \theta^{(2)}_{T,cz}) & var(\pi_c, \pi_c) \end{pmatrix}. \tag{5.9.40}$$

Thus the covariance matrix of π_c and $\theta_{T,cz}$ and the covariance of these products $cov(\pi_c S_{c1}(t), \pi_c S_{c0}(t))$ can be written as:

$$\begin{aligned} &Cov\left(\pi_c S_{c1}(t;\theta_{T,c1}), \pi_c S_{c0}(t;\theta_{T,c0})\right) \\ &= \left\{ \begin{array}{l} \left[\pi_c \frac{\partial S_{c1}(t;\theta_{T,c1})}{\partial \theta^{(1)}_{T,c1}},\ \pi_c \frac{\partial S_{c1}(t;\theta_{T,c1})}{\partial \theta^{(2)}_{T,c1}}, S_{c1}(t)\right] \Sigma_4 \\ \left[\pi_c \frac{\partial S_{c0}(t;\theta_{T,c0})}{\partial \theta^{(1)}_{T,c0}},\ \pi_c \frac{\partial S_{c0}(t;\theta_{T,c0})}{\partial \theta^{(2)}_{T,c0}}, S_{c0}(t)\right]^T \end{array} \right\} \end{aligned} \tag{5.9.41}$$

where Σ_4, the covariance matrix of relevant parameters, is given by:

$$\Sigma_4 = \begin{pmatrix} cov(\theta^{(1)}_{T,c1}, \theta^{(1)}_{T,c0}) & cov(\theta^{(1)}_{T,c1}, \theta^{(2)}_{T,c0}) & cov(\theta^{(1)}_{T,c1}, \pi_c) \\ cov(\theta^{(2)}_{T,c1}, \theta^{(1)}_{T,c0}) & cov(\theta^{(2)}_{T,c1}, \theta^{(2)}_{T,c0}) & cov(\theta^{(2)}_{T,c1}, \pi_c) \\ cov(\pi_c, \theta^{(1)}_{T,c0}) & cov(\pi_c, \theta^{(2)}_{T,c0}) & var(\pi_c, \pi_c) \end{pmatrix}. \tag{5.9.42}$$

Therefore, (5.9.38) can be re-written as:

$$
\begin{aligned}
&Var(ACE(t)) \\
&= \left\{ \begin{array}{l}
\left[\pi_c \frac{\partial S_{c1}(t;\theta_{T,c1})}{\partial \theta^{(1)}_{T,c1}},\ \pi_c \frac{\partial S_{c1}(t;\theta_{T,c1})}{\partial \theta^{(2)}_{T,c1}},\ S_{c1}(t;\theta_{T,c1})\right] \Sigma_{31} \left[\pi_c \frac{\partial S_{c1}(t;\theta_{T,c1})}{\partial \theta^{(1)}_{T,c1}},\ \pi_c \frac{\partial S_{c1}(t;\theta_{T,c1})}{\partial \theta^{(2)}_{T,c1}},\ S_{c1}(t;\theta_{T,c1})\right]^T \\
+\left[\pi_c \frac{\partial S_{c0}(t;\theta_{T,c0})}{\partial \theta^{(1)}_{T,c0}},\ \pi_c \frac{\partial S_{c0}(t;\theta_{T,c0})}{\partial \theta^{(2)}_{T,c0}},\ S_{c0}(t;\theta_{T,c0})\right] \Sigma_{30} \left[\pi_c \frac{\partial S_{c0}(t|\theta_{T,c0})}{\partial \theta^{(1)}_{T,c0}},\ \pi_c \frac{\partial S_{c0}(t;\theta_{T,c0})}{\partial \theta^{(2)}_{T,c0}},\ S_{c0}(t;\theta_{T,c0})\right]^T \\
-2\left[\pi_c \frac{\partial S_{c1}(t;\theta_{T,c1})}{\partial \theta^{(1)}_{T,c1}},\ \pi_c \frac{\partial S_{c1}(t;\theta_{T,c1})}{\partial \theta^{(2)}_{T,c1}}, S_{c1}(t;\theta_{T,c1})\right] \Sigma_4 \left[\pi_c \frac{\partial S_{c0}(t;\theta_{T,c0})}{\partial \theta^{(1)}_{T,c0}},\ \pi_c \frac{\partial S_{c0}(t;\theta_{T,c0})}{\partial \theta^{(2)}_{T,c0}}, S_{c0}(t;\theta_{T,c0})\right]^T
\end{array} \right\}
\end{aligned}
\tag{5.9.43}
$$

where Σ_{31} and Σ_{30} are given by equation (5.9.40) with $z = 1$ and 0, respectively. i.e.:

$$
\Sigma_{31} = \begin{pmatrix}
var(\theta^{(1)}_{T,c1}) & cov(\theta^{(1)}_{T,c1}, \theta^{(2)}_{T,c1}) & cov(\theta^{(1)}_{T,c1}, \pi_c) \\
cov(\theta^{(2)}_{T,c1}, \theta^{(1)}_{T,c1}) & var(\theta^{(2)}_{T,c1}) & cov(\theta^{(2)}_{T,c1}, \pi_c) \\
cov(\pi_c, \theta^{(1)}_{T,c1}) & cov(\pi_c, \theta^{(2)}_{T,c1}) & var(\pi_c, \pi_c)
\end{pmatrix} \tag{5.9.44}
$$

and

$$
\Sigma_{30} = \begin{pmatrix}
var(\theta^{(1)}_{T,c0}) & cov(\theta^{(1)}_{T,c0}, \theta^{(2)}_{T,c0}) & cov(\theta^{(1)}_{T,c0}, \pi_c) \\
cov(\theta^{(2)}_{T,c1}, \theta^{(1)}_{T,c0}) & var(\theta^{(2)}_{T,c0}) & cov(\theta^{(2)}_{T,c0}, \pi_c) \\
cov(\pi_c, \theta^{(1)}_{T,c0}) & cov(\pi_c, \theta^{(2)}_{T,c0}) & var(\pi_c, \pi_c)
\end{pmatrix}. \tag{5.9.45}
$$

Recall that Σ_4 is given in (5.9.42). All these covariance matrices are then available for our PPO-survival models. Consequently, it is straightforward to obtain the standard error of $ACE(t)$ by taking the square root of its variance $Var(ACE(t)$, that is:

$$
SE_{ACE(t)} = \sqrt{Var(ACE(t))}. \tag{5.9.46}
$$

5.10 Ignorable censoring mechanism PPO-survival model (I)

When the censoring mechanism is ignorable, as in the scenario when censoring time and failure time are entirely independent (Section 3.1.4) or when the non-informative censoring assumption is satisfied (Loeys & Goetghebeur, 2003), the PPO-survival model can then be simplified to the ignorable parametric potential-outcome (PPO) survival model, namely the ignorable PPO-survival model. Note that the assumption of *no censoring prior to the end of the follow-up* leads to the fact that no censoring distribution exist, as censoring times are a constant rather than a random variable (Section 4.3.1).

This implies that under the latter assumption, even though a constant censoring time does not mean its censoring mechanism is ignorable, we can still view it as a special case of the ignorable PPO-survival model, because no censoring distribution is involved.

So we summary the assumptions required for the ignorable PPO-survival model here:

5.10.1 Assumptions

1. The stable unit treatment value assumption (SUTVA) (Rubin, 1986; Imbens & Rubin, 1997; Peng et al., 2004): this assumes that the potential outcomes of each individual are unrelated to the treatment status of other individuals.

2. Random assignment: this assumes that each individual has been assigned randomly to a treatment group.

3. Exclusion restriction: this assumes that the assigned treatment has an effect on the outcome only through its effect on the received treatment.

4. Monotonicity: the actual received treatment $D_i(z)$ is a monotone function with respect to the potential treatment levels, 1 for being in the treatment group and 0 for being in the control group.

5. Non-informative censoring[3]: the failure times and censoring times are independent.

5.10.2 The likelihood function for the complete data

Under the assumptions given above, the ignorable PPO-survival model can be set up by following the logic of deriving the non-ignorable PPO-survival model. However, after obtaining the non-ignorable PPO-survival model, the relevant likelihood function can be easily obtained from the formulation of the likelihood structure of the non-ignorable PPO-survival model. In short we just need to remove the parts that are associated with censoring time C_i from equation (5.5.9). We then have:

$$\begin{aligned} & L_{i\in\{i:Z_i=z,D_i(z)=d\}}(\boldsymbol{\beta};u) \\ & \qquad = \left[f^i_u(z,t;\theta_{T,uz})\right]^{R_i}\left[S^i_{uz}(v,t;\theta_{T,uz})\right]^{1-R_i}\pi_{u,zd} \\ & \qquad = l^i_\theta(\boldsymbol{\theta},u)l^i_\pi(\boldsymbol{\pi},u). \end{aligned} \tag{5.10.1}$$

Therefore, for the ignorable PPO-survival model, the likelihood formulation with respect to θ for complete data, can then be expressed by equation (5.10.2) below:

$$\begin{aligned} L^F(\boldsymbol{\beta}) = L(\boldsymbol{\beta},u) &= \prod_i L_i(\boldsymbol{\beta},u) \\ &= \begin{cases} \prod_{i\in\{i:U_i=c,Z_i=1,D_i(1)=1\}}\mathcal{F}^i_{c,11}(\theta_{T,c1})\pi_{c,11} \\ \prod_{i\in\{i:U_i=a,Z_i=1,D_i(1)=1\}}\mathcal{F}^i_{a,11}(\theta_{T,a1})\pi_{a,11} \\ \prod_{i\in\{i:U_i=c,Z_i=0,D(0)=0\}}\mathcal{F}^i_{c,00}(\theta_{T,c0})\pi_{c,00} \\ \prod_{i\in\{i:U_i=n,Z_i=0,D_i(0)=0\}}\mathcal{F}^i_{n,00}(\theta_{T,n0})\pi_{n,00} \\ \prod_{i\in\{i:U_i=n,Z_i=1,D(0)=0\}}\mathcal{F}^i_{n,10}(\theta_{T,n0})\pi_{n,10} \\ \prod_{i\in\{i:U_i=a,Z_i=0,D(0)=1\}}\mathcal{F}^i_{a,01}(\theta_{T,a1})\pi_{a,01}; \end{cases} \end{aligned} \tag{5.10.2}$$

[3]For our ignorable PPO-survival model, the *non-informative censoring* assumption can be replaced by the assumption of *Censored at the end of the follow-up only*, or *No censoring prior the end of the follow-up* or *No loss to follow-up*, as under these two assumptions, the model formulae are same.

where $\mathcal{F}^i_{u,zd}(\theta_{T,uz})$ is given by the following formulation:

$$\mathcal{F}^i_{u,zd}(\theta_{T,uz}) = \left[f^i_u(v,t;\theta_{T,uz})\right]^{R_i}\left[S^i_u(v,t;\theta_{T,uz})\right]^{1-R_i}. \tag{5.10.3}$$

Note that this is dissimilar to equation (5.5.12), which pertains to the more complex non-ignorable PPO-survival model.

5.10.3 The likelihood function for the incomplete data

As stated in Section 5.5.3, compliance type U_i can be viewed as a variable with missing data. Thus we use summation, rather than integration in equation (5.10.2) with respect to compliance type $U_i = u$, as it is a variable with up to three categories ($u \in \{c, n, a\}$). We then have the form of the likelihood function for the incomplete data for the ignorable PPO-survival model. This is:

$$\begin{aligned} L(\boldsymbol{\beta}) &= \sum_{u\in\{c,n,a\}} L(\boldsymbol{\beta}, u) \\ &= \begin{cases} \prod_{i\in\mathcal{A}(1,1)} \left\{\left(\mathcal{F}^i_{c,11}(\theta_{T,c1})\pi_{c,11}\right) + \left(\mathcal{F}^i_{a,11}(\theta_{T,a1})\pi_{a,11}\right)\right\} \\ \prod_{i\in\mathcal{A}(0,0)} \left\{\left(\mathcal{F}^i_{c,00}(\theta_{T,c0})\pi_{c,00}\right) + \left(\mathcal{F}^i_{n,00}(\theta_{T,n0})\pi_{n,00}\right)\right\} \\ \prod_{i\in\mathcal{A}(1,0)} \left(\mathcal{F}^i_{n,10}(\theta_{T,n0})\pi_{n,10}\right) \prod_{i\in\mathcal{A}(0,1)} \left(\mathcal{F}^i_{a,01}(\theta_{T,a1})\pi_{a,01}\right). \end{cases} \end{aligned} \tag{5.10.4}$$

5.10.4 The ignorable PPO-survival model parameter specification

Clearly, equations (5.10.2) and (5.10.4) show that $\boldsymbol{\theta} = \boldsymbol{\theta}_T$ in the ignorable PPO model. Similarly to the non-ignorable PPO-survival model, we can specify $\boldsymbol{\theta}_T$ as follows:

$$\boldsymbol{\theta} = \boldsymbol{\theta}_T = (\theta_{T,c1}, \theta_{T,c0}, \theta_{T,n}, \theta_{T,a}). \tag{5.10.5}$$

Since only the failure time distribution is included in the likelihood function (equation (5.10.1)), we need only to specify one standard survival distribution for the ignorable PPO-survival model. Our convention is as follows: if the Weibull model is chosen, the ignorable parametric potential-outcome (PPO) model is called the ignorable PPO-survival Weibull model (called the

IGN-W model for short). The IGN-W model is thus a parametric potential-outcome (PPO) survival model for noncompliance with an ignorable censoring mechanism. Similarly, if the specified distribution is the log-normal, our resultant model is the ignorable PPO-survival log-normal model, called the IGN-L model. Table 5.6 gives the notation for these ignorable PPO-survival models.

Table 5.6: Specification of the ignorable PPO-survival models.

Distribution	Ignorable survival model
Weibull	IGN-W model
log-normal	IGN-L model

For these two ignorable PPO-survival models, the pertinent parameters are only those associated with the failure time distributions, as no censoring time distribution is involved. We then have the parameters shown in Table 5.7.

Table 5.7: Ignorable PPO-survival model parameter vectors.

Model name	model parameters
IGN-W	$\rho_{T,c1},\ \rho_{T,c0},\ \rho_{T,n},\ \rho_{T,a},\ \alpha_T$
IGN-L	$\mu_{T,c1},\ \mu_{T,c0},\ \mu_{T,n},\ \mu_{T,a},\ \sigma_T$

5.10.5 The posterior probability of compliance type

Note that in the ignorable PPO-survival model no censoring distribution is involved. In this case, then the posterior probabilities of the two specific cases (the case of $Z_i = 1$ and $D_i(1) = 1$, and the case of $Z_i = 0$ and $D_i(0) = 0$)

can be simplified as follows:

$$Pr(U_i = u|(T_i(1), C_i(1))^{obs},\ Z_i = 1,\ D_i = 1;\ \boldsymbol{\theta}^k) = \begin{cases} \frac{f_u(1,t;\theta_{T,u1})\pi_{u,11}}{f_c(1,t;\theta_{T,c1})\pi_{c,11}+f_a(1,t;\theta_{T,a1})\pi_{a,11}} & \text{if } R_i = 1,\ T_i(1) = t, \\ \frac{S_u(1,t;\theta_{T,u1})\pi_{u,11}}{S_{c1}(t;\theta_{T,c1})\pi_{c,11}+S_a(1,t;\theta_{T,a1})\pi_{a,11}} & \text{if } R_i = 0, C_i(1) = t; \end{cases} \tag{5.10.6}$$

for $u \in \{c,\ a\}$ and

$$Pr(U_i = u|(T_i(1, C_i(1))^{obs},\ Z_i = 0,\ D_i = 0;\ \boldsymbol{\theta}_0) = \begin{cases} \frac{f_u(0,t;\theta_{T,u0})\pi_{u,00}}{f_c(0,t;\theta_{T,c0})\pi_{c,00}+f_n(0,t;\theta_{T,n})\pi_{n,00}} & \text{if } R_i = 1,\ T_i(0) = t, \\ \frac{S_u(0,t;\theta_{T,u0})\pi_{u,00}}{S_{c0}(t;\theta_{T,c0})\pi_{c,00}+S_n(0,t;\theta_{T,n})\pi_{n,00}} & \text{if } R_i = 0, C_i(0) = t; \end{cases} \tag{5.10.7}$$

for $u \in \{c,\ n\}$.

Therefore, for the ignorable PPO-survival model, P^i_{ac}, P^i_{nc}, Q^i_{ac}, Q^i_{nc} have the following equivalences:

$$\begin{aligned} P^i_{ac} &= \frac{f_c(1,t;\theta_{T,c1})\pi_{c,11}}{f_c(1,t;\theta_{T,c1})\pi_{c,11} + f_a(1,t;\theta_{T,a1})\pi_{a,11}} && \text{for } i \in \mathcal{C}(1,1,1),\ T_i(1) = t\ , \\ P^i_{nc} &= \frac{f_c(0,t;\theta_{T,c0})\pi_{c,00}}{f_c(0,t;\theta_{T,c0})\pi_{c,00} + f_n(0,t;\theta_{T,n})\pi_{n,00}} && \text{for } i \in \mathcal{C}(0,0,1), T_i(0) = t, \\ Q^i_{ac} &= \frac{S_c(1,t;\theta_{T,c1})\pi_{c,11}}{S_c(1,t;\theta_{T,c1})\pi_{c,11} + S_a(1,t;\theta_{T,a1})\pi_{a,11}} && \text{for } i \in \mathcal{C}(1,1,0), C_i(1) = t, \\ Q^i_{nc} &= \frac{S_c(0,t;\theta_{T,c0})\pi_{c,00}}{S_c(0,t;\theta_{T,c0})\pi_{c,00} + S_n(0,t;\theta_{T,n})\pi_{n,00}} && \text{for } i \in \mathcal{C}(0,0,0), C_i(0) = t. \end{aligned} \tag{5.10.8}$$

5.10.6 The expectation function $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^k)$

For the ignorable PPO-survival model, after substituting equations (5.10.8) and (5.9.11) into equation (5.9.12), $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^k)$ can then be specified as follows:

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^k) = \sum_i Q^i(\boldsymbol{\theta}, \boldsymbol{\theta}^k) =$$

$$\begin{cases} \sum_{i\in\mathcal{C}(1,1,1)} \{log\left(f_c^i(1,t;\theta_{T,c1})\pi_{c,11}\right) P_{ac}^i + log\left(f_a^i(1,t;\theta_{T,a})\pi_{a,11}\right)(1-P_{ac}^i)\} \\ \sum_{i\in\mathcal{C}(1,1,0)} \{log\left(S_c^i(1,t;\theta_{T,c1})\pi_{c,11}\right) Q_{ac}^i + log\left(S_a^i(1,t;\theta_{T,a})\pi_{a,11}\right)(1-Q_{ac}^i)\} \\ \sum_{i\in\mathcal{C}(0,0,1)} \{log\left(f_c^i(0,t;\theta_{T,c0})\pi_{c,00}\right) P_{nc}^i + log\left(f_n^i(0,t;\theta_{T,n})\pi_{n,00}\right)(1-P_{nc}^i)\} \\ \sum_{i\in\mathcal{C}(0,0,0)} \{log\left(S_c^i(0,t;\theta_{T,c0})\pi_{c,00}\right) Q_{nc}^i + log\left(S_n^i(0,t;\theta_{T,n})\pi_{n,00}\right)(1-Q_{nc}^i)\} \\ \sum_{i\in\mathcal{C}(1,0,1)} log\left(f_n^i(1,t;\theta_{T,n})\pi_{n,10}\right) \sum_{i\in\mathcal{C}(1,0,0)} log\left(S_n^i(1,t;\theta_{T,n})\pi_{n,10}\right) \\ \sum_{i\in\mathcal{C}(0,1,1)} log\left(f_a^i(0,t;\theta_{T,a})\pi_{a,01}\right) \sum_{i\in\mathcal{C}(0,1,0)} log\left(S_a^i(0,t;\theta_{T,a})\pi_{a,01}\right). \end{cases} \tag{5.10.9}$$

5.10.7 Computation

Following the same process as stated in Section 5.9, the ignorable PPO-survival model can be implemented directly. This means that in the special case (when the censoring mechanism is ignored), we can still use all the methodologies given in Section 5.9. For example, to maximise the expectation function for this model $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^k)$ (given in equation (5.10.9)), we follow Section 5.9.3; all formulations given in Section 5.9.4 can also be used to calculate the standard error with respect to MLEs obtained in the M-step and formulations with respect to the estimated $\widehat{CACE}(t)$ and $\widehat{ACE}(t)$ (see 5.9.4).

5.11 Summary

When the noncompliance issue has to be accounted for in likelihood-based survival models, it is most likely that an underlying non-ignorable censoring mechanism exists. Indeed, as shown in equation (5.5.16), the censoring time distribution must be included in the derivation of the likelihood function unless censoring is ignorable (Loeys & Goetghebeur, 2003). For such cases, we formulated a non-ignorable censoring mechanism parametric

potential-outcome (PPO) survival model, denoted the non-ignorable PPO-survival model for short. More specifically, when failure and censoring times are assumed to follow a Weibull or log-normal distribution, we have derived and coined the NIGN-WW and the NIGN-LL models, respectively. Another model, the so-called ignorable parametric potential-outcome (PPO) survival model, (the ignorable PPO-survival model for short), is also given in this chapter. The ignorable PPO-survival model pertains to the special case where no censoring occurs prior to the end of follow-up or when the non-informative censoring assumption holds (Loeys & Goetghebeur, 2003). Similarly, the IGN-W and the IGN-L model are ignorable PPO-survival models for the assumed scenarios where the failure times follow Weibull or log-normal distributions.

Because compliance type is an unobservable variable, we adopt the EM algorithm to maximise the likelihood function for the complete data, where the compliance type is considered to be an observable variable with missing data. We adapt Louis's method (Louis, 1982) for the PPO-survival models (including both non-ignorable and ignorable censoring mechanism), to evaluate the standard error associated with the MLEs, obtained via the EM algorithm. However, the Bootstrap approach (Efron & Tibshirani, 1993) is necessary in this scenario to estimate the standard error of parameter $\boldsymbol{\pi}$ as it is not involved in the EM algorithm.

Theorem 1 provides a valuable property of AFT model, which holds for both Weibull and log-normal distributions. Thereby, the accelerated failure time (AFT) model guarantees that each subgroup has the same shape parameter, for the Weibull model, or variance parameter for the logarithm of failure time (or the logarithm of censoring time) for the log-normal model.

Our approach achieves a reduction in the number of parameters for the non-ignorable and ignorable survival models, whether we use the NIGN-WW (or the NIGN-LL) model or the IGN-W (or the IGN-L) model. This leads to reduced computational costs and time, of particular value to censored RCT data applications with noncompliance.

Chapter 6

HIP breast cancer RCT application

The parametric potential-outcome (PPO) survival models, introduced in Chapter 5, which include both non-ignorable and ignorable censoring mechanisms for noncompliance analysis of survival outcomes, have been established theoretically. The survival outcome distributions have failure and censoring times specified by the Weibull or the log-normal distribution. This chapter demonstrates how these non-ignorable and ignorable survival models perform on an analysis of the HIP breast cancer data (Shapiro et al., 1988; Baker, 1998). Furthermore, the Frangakis-Rubin (F-R) method (Frangakis & Rubin, 1999) for considering the noncompliance and censoring (non-response) mechanism jointly from non-parametric perspective (Section 2.3.1), is also applied to the HIP data. All programmes used in this thesis were run using in-house code for MATLAB (7.4.0, R2007a) software (Higham & Higham, 2000) (www.mathworks.com/products/matlab).

The structure of this chapter is as follows: Section 6.1 provides basic information on the Health Insurance Plan (HIP) of the Greater New York breast cancer RCT (Shapiro et al., 1988). Sections 6.2 and 6.3 show the application of our models to the HIP data. The Frangakis-Rubin (F-R) model is briefly introduced in Section 6.4. The results of our analysis of the HIP data using the F-R model are also given in Section 6.4. Section 6.5 provides comparisons of the results obtained via the non-parametric (F-R) model and our parametric potential-outcome (PPO) survival models.

6.1 The HIP study

6.1.1 HIP breast cancer programme

The breast cancer data used in this chapter were collected from a breast screening study conducted in the USA by the Health Insurance Plan (HIP) of Greater New York during the 1960s (Shapiro et al., 1988). Participants were women who enrolled in the HIP for one year or more when the programme started in 1963. These women were asymptomatic of breast cancer and aged between 40 and 64 years old at entry into the HIP trial.

A total of 60,696 eligible women were included in this programme, and all were randomly assigned into either the *screening* or the *control* group. Women in the *screening* group were offered free annual breast examinations four times. These examinations consisted of film mammography (cephalocaudal and later views of each breast), a clinical examination of each breast by a physician (usually a surgeon), and an interview for demographic and other background information including a health history (Shapiro et al., 1988). The women in the so-called *control group* received only their usual health care (Shapiro et al., 1988). All participants in this study were followed up in the long-term for up to 18 years from their date of entry. The event of interest was death from breast cancer, and the outcome of interest was the observed time (in years), for each individual from entry to death from breast cancer or to loss to follow-up. Clearly, for those women diagnosed with breast cancer in the study, there was no loss to follow up. In previous related research (Shapiro et al., 1988; Baker, 1998), the mortality that due to breast cancer, was the main outcome variable of interest and was used to evaluate the efficacy of the breast screening programme.

Approximately one-third of the women assigned to the *screening group* refused the screening examination. Thus the HIP trial satisfies the special case of noncompliance, known as *all-or-none compliance* (Baker, 1997), wherein a switch from the assigned treatment to another treatment occurs immediately after randomisation. Women who were allocated to the *screening* group, but refused to take the examination, are viewed as being switched to the control group directly at the beginning of the trial. Hence the HIP breast cancer trial is considered to be a RCT with *all-or-none compliance* (Baker, 1997).

In addition, the HIP data satisfies the assumption of *no access to treatment* for the control group. This assumption has also been made by Frangakis & Rubin (1999) and by Loeys & Goetghebeur (2003). This is considered to be true in the HIP trial, as no women in the *control* group received the screening examination.

Previous analyses of the HIP breast cancer data (Shapiro, 1977; Shapiro et al., 1982, 1988) were mostly based on the intention-to-treat (ITT) principle, and focused on estimating the cumulative mortality rate of breast cancer and the survival rate, over particular fixed follow-up durations, rather than on estimating survival or hazard functions. From a noncompliance perspective, only Baker (1998) reported the estimated cost-effectiveness of the HIP intervention.

As was noted in Chapter 1, *noncompliance* presents a challenge to traditional statistical methods for the analysis of RCTs. This is because neither an analysis which ignores non-compliers, nor an analysis which focuses on treatment (actually received), can provide precise and reliable answers (O'Malley & Normand, 2005; Mealli et al., 2004; Loeys & Goetghebeur, 2003). In this chapter, we aim to make valid inferences regarding the effect of mammography on breast cancer mortality. In the former case, the observed outcomes of the intervention group are a mixture of outcomes for women receiving and not receiving the intervention. In the latter case, the two groups are not necessarily comparable, because of the self-selection of treatment for the intervention group means that treatment received is not randomized. Consequently, treatment groups may differ on so-called background factors related to breast cancer and also differ on factors related to breast cancer mortality risk.

6.1.2 Pre-processing of the HIP breast cancer data

The original HIP data (Shapiro et al., 1988; Baker, 1998) includes 60,696 observations, in which the enrolment date and the last follow-up date for all individuals had been recorded, as well as their survival status. The latter indicates whether an individual died of breast cancer during the observation period or not. Note that in the HIP breast cancer trial, there were two

examiners independently took the responsibility of determining the course of each single case of death. For those individuals who died but not due to breast cancer, they were treated as censored individuals in the trial (Shapiro et al., 1988).

The observed time (from entry to death or last follow-up) was given in days in the original HIP data. In this thesis, the survival status of an individual is the so-called censoring indicator (denoted by $R = 1$ for death from breast cancer, 0 otherwise) and observed times are counted in years, as in Baker (1998), which was obtained by the following formula:

$$\text{observed time (in years)} = \frac{\text{days for entry to death or last follow-up}}{365}. \tag{6.1.1}$$

We note that 164 individuals in the original HIP data have the same date for their enrolment and their last follow-up date. This implies that these individuals did not, in fact, take part in the HIP study and thus should not be counted as "valid" individuals in the HIP trial. Therefore, in this thesis, we focus on analyzing the subset of the HIP data (60,696 - 160 = 60,532 women) rather than the original HIP data.

The HIP_{18} dataset: The first data subset studied follows Shapiro's report on the HIP study (Shapiro et al., 1988), wherein the maximum observed time is fixed at 18 years. This is because Shapiro et al. (1988) purports that in the HIP study, the maximum follow-up period is 18 years. Thus, women whose observed times are greater than 18 years (in the original HIP data set) are all preset at 18. Their censoring indicators are given as 0, as they can be viewed as censored at year 18 (after entry).

The HIP_7 dataset: The second data subset is obtained by following the arguments of Baker (1998). Baker mentions that: "by seven years, the number of breast cancers in the control and screen groups have equalized, so presumably they represent all breast cancers that could have benefitted from screening." To reduce the variance of estimators, Baker (1998) focused only on analyzing those individuals with breast cancer diagnosed in their first seven years of follow-up. Baker also points out that the choice of seven years

(for the HIP study) was conservative; other authors have used earlier times of so-called equalisation (Baker, 1998).

As with Baker (1998), we also chose to analyse the HIP data with seven years as the maximum follow-up period. The benefit from the screening examination should therefore be considered "realised" in seven rather than in eighteen years (Baker, 1998). Note that, as we are interested in investigating how the screening examination impacts the longevity of women who are in general good health (rather than breast cancer patients). Hence, unlike Baker (1998), in this thesis, we focus on analysing all women in the HIP trial rather than those who had been diagnosed with breast cancer during the observation period.

After truncating the observed time at 7 years, individuals whose observed times were greater than seven years have been given 0 as their censored indicators value, which shows that these individuals had not died at that time (7 years). Therefore, in this HIP dataset, denoted by HIP_7, the maximum observed time is 7 years; individuals whose observed time are greater than 7 years (in the original HIP data) have $R_i = 0$. This indicates that individuals whose observed time has been truncated (in the HIP_7 data) did not die from breast cancer during the the follow-up period (7 years after entry). Summary statistics for the three HIP datasets (the original HIP, HIP_{18} and HIP_7) are given in Table 6.1 (on page 135).

As stated before, we focused on analysing the two aforementioned HIP data subsets, namely HIP_{18} and HIP_7, with our NIGN-WW and NIGN-LL survival models. The results are given in Section 6.2. Furthermore, for the purpose of comparison, these two HIP data subsets are also re-analysed by the Frangakis-Rubin (F-R) model (Frangakis & Rubin, 1999). To date, the F-R model has not yet been applied to any time-to-event data in the literature. The results of the re-analysis of the HIP data subsets are given in Section 6.4. In addition, a specific IGN-survival model, the so-called IGN-W model, is applied to the HIP_{18} and the HIP_7 datasets (Section 6.3).

Table 6.1: Summary statistics of HIP data sets

Original HIP data

	Total	Screen invited $Z=1$	Control $Z=0$	Screen received $Z=1, D=1$	Screen refused $Z=1, D=0$	No Screen received $D=0$
Sample size N	60,696	30,131	30,565	20,147	9,984	40,549
Mean of observed time $\bar{Y}_i$	14.051	14.1578	13.9457	14.905	12.6499	13.6266
Proportion of event $\bar{R}_i$	0.0129	0.0123	0.0135	0.0123	0.0125	0.0133
Proportion of screen invited $\bar{Z}_i$	0.4964	1	0	1	1	0.2462
Proportion of screen received $\bar{D}_i$	0.3319	0.6686	0	1	0	0

HIP_{18} data

	Total	Screen invited $Z=1$	Control $Z=0$	Screen received $Z=1, D=1$	Screen refused $Z=1, D=0$	No Screen received $D=0$
Sample size N	60,532	30,058	30,474	20,132	9,926	40,400
Mean of observed time $\bar{Y}_i$	13.9756	14.0816	13.8709	14.8112	12.6018	13.5591
Proportion of event $\bar{R}_i$	0.0126	0.0119	0.0134	0.0119	0.0119	0.0130
Proportion of screen invited $\bar{Z}_i$	0.4966	1	0	1	1	0.2457
Proportion of screen received $\bar{D}_i$	0.3326	0.6698	0	1	0	0

HIP_7 data

	Total	Screen invited $Z=1$	Control $Z=0$	Screen received $Z=1, D=1$	Screen refused $Z=1, D=0$	No Screen received $D=0$
Sample size N	60,532	30,058	30,474	20,132	9,926	4,0400
Mean of observed time $\bar{Y}_i$	6.5503	6.5813	6.5197	6.7393	6.2607	6.4561
Proportion of event $\bar{R}_i$	0.0030	0.0021	0.0038	0.0020	0.0024	0.0035
Proportion of screen invited $\bar{Z}_i$	0.4966	1.0000	0	1.0000	1.0000	0.2457
Proportion of screen received $\bar{D}_i$	0.3326	0.6698	0	1	0	0

$Z_i = 1$: invitation for screening; $D_i = 1$: receipt of screening and $R_i = 1$: death from breast cancer.

6.2 HIP data analysis: via the non-ignorable censoring mechanism PPO-survival models

For the HIP trial, it is assumed that whether a participant died from breast cancer or not would not be related to any other participant's treatment status. We can then suppose Assumption 1 (given in Section 5.2), the SUTVA assumption is satisfied. Assumption 2 (random assignment) clearly holds as all participants in the HIP trial were randomly assigned to the *screening* or the *control* group (Shapiro et al., 1988). Some participants assigned to the screening group refused to take the screening examination (Shapiro et al., 1988; Baker, 1998). This implies that these individuals did not follow their allocated randomised assigned treatment (Z_i). Assumption 3, namely that of *exclusion restriction*, can then be assumed. This means that the assigned treatment (Z_i) affects the outcome of interest only through the actual received treatment ($D_i(Z_i)$). We also suppose Assumption 4, that of *monotonicity*, holds in the HIP trial, as it is more acceptable to suppose that no defiers exist than to have defiers in a RCT (Angrist et al., 1996; Frangakis & Rubin, 1999; Baker, 1998; Yau & Little, 2001; Peng et al., 2004; O'Malley & Normand, 2005; Mattei & Mealli, 2007). Finally, we suppose that the *latent ignorability* assumption holds in the HIP trial, i.e. that the potential failure time and censoring time are independent, conditional on the latent compliance type. Hence, all assumptions given in Section 5.2 are assumed satisfied in the HIP trial. In addition, for all individuals in this trial, there are only two possible types of compliance behavior, namely *compliers* or *never-takers*. Note that no woman assigned to the control group could receive the screening examination, which implies that no always-takers exist in the HIP breast screening trial. In other words, the assumption of *no access to treatment for the control* (Loeys & Goetghebeur, 2003; Frangakis & Rubin, 1999) holds in the HIP breast screening trial. However, it is not a necessary assumption for our PPO-survival models (derived in Chapter 5).

6.2.1 Model specification for the HIP trial

Since no always-takers exist in the HIP trial, the number of compliance types in the HIP data has thus been reduced to two, and the number of subgroups of

interest in the HIP data are thus reduced to three (Subgroup j; $j = 1, 2, 3$), specified as follows:

Subgroup1 : compliers ($U_i = c$) from the *screening* group ($Z_i = 1$);

Subgroup2 : compliers ($U_i = c$) from the *control* group ($Z_i = 0$);

Subgroup3 : never-takers ($U_i = n$) from either the *screening* ($Z_i = 1$) or the *control* ($Z_i = 0$) group.

Consequently, for the HIP breast cancer data, the model parameter vector $\boldsymbol{\beta}$, given in equation (5.5.23), is simplified as follows:

$$\begin{aligned} &\boldsymbol{\beta} = (\boldsymbol{\theta}, \boldsymbol{\pi}); \quad \text{where} \\ &\boldsymbol{\theta} = (\theta_{T,c1}, \theta_{T,c0}, \theta_{T,n}, \alpha_T, \theta_{C,c1}, \theta_{C,c0}, \theta_{C,n}, \alpha_C), \quad \text{and} \\ &\boldsymbol{\pi} = (\pi_{c,00}). \end{aligned} \tag{6.2.1}$$

Recall that only $\boldsymbol{\theta}$ need to be estimated using the EM algorithm, $\boldsymbol{\theta}$ is the interest of model parameter.

Note that the parameters corresponding to the always-takers subgroup in equation (5.5.23) are thus excluded in equation (6.2.1), because no always-takers exist in the HIP trial. This can be denoted by $\pi_a = 0$, which indicates that the *ith* individual in the HIP trial, has 0 (zero) probability of being an always-taker. Therefore, the parameter $\pi_{c,11}$, defined by the first equation in (5.5.27), is fixed as 1 (because $\pi_a = 0$), i.e. $\pi_{c,11} \equiv 1$. So for the HIP data, $\pi_{c,11}$ is clearly not a parameter that needs to be estimated. Furthermore, the parameter $\pi_{c,00}$, defined by the second equation in (5.5.27) is equivalent to π_c, ($\pi_{c,00} \equiv \pi_c$) because $\pi_c + \pi_n \equiv 1$ in the case of the HIP study.

Hence, for the NIGN-WW model, which assumes that both failure and censoring time follow a Weibull distribution, the interest model parameter vector, $\boldsymbol{\theta}_{WW}$, includes eight elements, i.e.:

$$\boldsymbol{\theta}_{WW} = (\rho_{T,c1},\ \rho_{T,c0},\ \rho_{T,n},\ \alpha_T,\ \rho_{C,c1},\ \rho_{C,c0},\ \rho_{C,n},\ \alpha_C)^T, \tag{6.2.2}$$

where $\rho_{c1},\ \rho_{c0}$ and ρ_n represent the scale parameters of three Weibull distributions for failure time (T_i) associated with the relevant subgroups (Subgroup1,

Subgroup2 and Subgroup3), with α_T as the common shape parameter for these Weibull distributions; and $\rho_{C,c1}$, $\rho_{C,c0}$ and $\rho_{C,n}$ represent the scale parameters of three Weibull distributions for censored time (C_i) associated with the same three relevant subgroups (Subgroup1, Subgroup2 and Subgroup3) and with α_c as the common shape parameter for these distributions.

When the log-normal is assumed to be the distribution for failure and censoring time , the interest NIGN-LL model parameter vector, denoted by $\boldsymbol{\theta_{LL}}$, is as follows:

$$\boldsymbol{\theta_{LL}} = (\mu_{T,c1},\ \mu_{T,c0},\ \mu_{T,n},\ \sigma_T,\ \mu_{C,c1},\ \mu_{C,c0},\ \mu_{C,n},\ \sigma_C), \tag{6.2.3}$$

where μ_{c1}, μ_{c0} and μ_n denote the mean of the logarithm of failure time, corresponding to the three extant subgroups (Subgroup1, Subgroup2 and Subgroup3); σ_T is their common variance (of the logarithm of failure time); and $\mu_{C,c1}$, $\mu_{C,c0}$, $\mu_{C,n}$ and σ_C represent the parameters corresponding to the logarithm of censoring time.

The model parameters corresponding to the specific non-ignorable PPO-survival models (for the HIP trial) are given in Table 6.2.

Table 6.2: Non-ignorable PPO-survival model parameters for the HIP trial.

Model name	model parameter vector
NIGN-WW	$(\rho_{T,c1},\ \rho_{T,c0},\ \rho_{T,n},\ \alpha_T,\ \rho_{C,c1},\ \rho_{C,c0}, \rho_{C,n},\ \alpha_C)$
NIGN-LL	$(\mu_{T,c1},\ \mu_{T,c0},\ \mu_{T,n},\ \sigma_T,\ \mu_{C,c1},\ \mu_{C,c0}, \mu_{C,n},\ \sigma_C)$

The estimated values of the interest model parameter vectors $\boldsymbol{\theta}$, associated with the two non-ignorable survival models, namely the NIGN-WW and NIGN-LL models, are given in Appendix C. Specifically, Tables C-1 and C-2 (on page 249) give analogous estimates for the HIP_{18} data; Tables C-3 and C-4 (on page 250), for the HIP_7 data. Details on the iteration process used in the EM algorithm (Dempster et al., 1977) for these two specific non-ignorable PPO-survival models are detailed in Appendix B, where Tables C-8 (on page 252) and C-9 (on page 253) give the iteration of convergence in the EM algorithm for the HIP_{18} dataset, and Tables C-10 (on page 254) and C-11 (on page 255), for the HIP_7 dataset.

6.2.2 Estimation of causal effects: CACE(t) and ACE(t)

Table 6.3: Estimated $CACE(t)$ and associated standard errors per model.

Models	NIGN-WW model	NIGN-LL model
Data	HIP_{18}	
Years	CACE (SE_{CACE})	CACE (SE_{CACE})
1	0.19E-4 (0.10E-4)	0.15E-4 (0.07E-4)
5	3.56E-4 (1.86E-4)	5.22E-4 (2.19E-4)
10	12.6E-4 (6.57E-4)	17.9E-4 (7.17E-4)
15	26.2E-4 (13.7E-4)	32.1E-4 (13.2E-4)
Data	HIP_7	
Years	CACE (SE_{CACE})	CACE (SE_{CACE})
1	0.66E-4 (0.24E-4)	0.62E-4 (0.25E-4)
5	14.8E-4 (3.83E-4)	15.7E-4 (4.02E-4)

For causal inference in time-to-event studies, we focus on estimating the causal effects in terms of CACE(t) and ACE(t). These were previously defined by equations (3.3.7) and (3.3.8) in Chapter 3. Tables 6.3 (on page 139) and 6.4 (on page 140) give estimated CACE(t) and ACE(t) for the two non-ignorable PPO survival models (the NIGN-WW and the NIGN-LL model). Their associated standard errors are also given in Tables 6.3 and 6.4, which are calculated by following the formulation given in equations (5.9.37) and (5.9.46).

Since CACE(t) and ACE(t) both are the functions of the MLE and appealing to large-sample normality of the MLE, the 95% lower and upper confidence limits (LCL and UCL) of CACE(t) are calculated easily as follows:

$$\begin{aligned} LCL_{cace}(t) &= \widehat{CACE}(t) - 1.96 SE_{\widehat{cace}}(t) \\ UCL_{cace}(t) &= \widehat{CACE}(t) + 1.96 SE_{\widehat{cace}}(t). \end{aligned} \tag{6.2.4}$$

Similarly, the 95% lower and the upper confidence limits (LCL and UCL)

Table 6.4: Estimated $ACE(t)$ and associated standard errors per model.

Models	$NIGN-WW$ model	$NIGN-LL$ model
Data	HIP_{18}	
Years	ACE (SE_{ACE})	ACE (SE_{ACE})
1	0.13E-4 (0.07E-4)	0.10E-4 (0.05E-4)
5	2.38E-4 (1.25E-4)	3.49E-4 (1.47E-4)
10	8.41E-4 (4.40E-4)	11.6E-4 (4.80E-4)
15	17.5E-4 (9.19E-4)	21.5E-4 (8.81E-4)
Data	HIP_7	
Years	ACE (SE_{ACE})	ACE (SE_{ACE})
1	0.44E-4 (0.16E-4)	0.41E-4 (0.16E-4)
5	9.90E-4 (2.56E-4)	10.5E-4 (2.69E-4)

of ACE(t) are calculated via the following formulae:

$$\begin{aligned} LCL_{ace}(t) &= \widehat{ACE}(t) - 1.96 SE_{\widehat{ace}}(t) \\ UCL_{ace}(t) &= \widehat{ACE}(t) + 1.96 SE_{\widehat{ace}}(t). \end{aligned} \tag{6.2.5}$$

The associated Z values of CACE(t) and ACE(t) are also given for all analyses. The Z value is calculated by using the following equations:

$$Z - value_{cace}(t) = \frac{\widehat{CACE}(t) - 0}{SE_{\widehat{cace}}(t)} \tag{6.2.6}$$

and

$$Z - value_{ace}(t) = \frac{\widehat{ACE}(t) - 0}{SE_{\widehat{ace}}(t)}; \tag{6.2.7}$$

where 0 is the assumed under the null mean value of CACE(t) or ACE(t). The p-value derives from $1 - \Phi(Z - value)$ (one tail), where $\Phi()$ denotes the standard normal cumulative probability function.

Further details regarding the estimated ACE(t) and CACE(t), obtained via our non-ignorable PPO-survival models for the HIP_{18} and HIP_7 data sets are given in Appendix C, where Tables C-13, C-14 give the estimated ACE(t) for the HIP_{18} dataset via the NIGN-WW and NIGN-LL model,

respectively, and Tables C-16 and C-17 give the estimated CACE(t) for the HIP_{18} data set via the same two models. Tables C-20, C-21, C-22 and C-23 are analogous for the analysis of the HIP_7 data set.

Based on these estimated values of ACE(t) and CACE(t), the curves of the estimated CACE(t) and ACE(t) against survival time (years survival post entry) are plotted in the following figures per data subset analysed. For the the HIP_{18} data, Figures 6.1 and 6.2 (on page 142) show the estimated CACE(t) curves obtained using the NIGN-WW and NIGN-LL model, respectively, and their associated ACE(t) curves are given in Figures 6.3 and 6.4 (on page 143); the corresponding curves of estimated CACE(t) and ACE(t) (with the same names) for the HIP_7 data are given in Figures 6.5 and 6.6 (on page 144) and Figures 6.7 and 6.8 (on page 145).

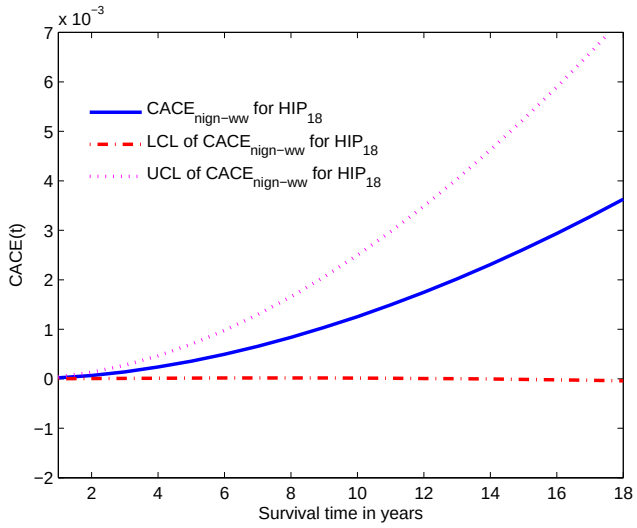

Figure 6.1: Estimated CACE(t) curve using the NIGN-WW model with 95% confidence interval: HIP_{18} data. The dash-dot line and the dotted line indicate the Lower Confidence Limit(LCL) and the Upper Confidence Limit (UCL).

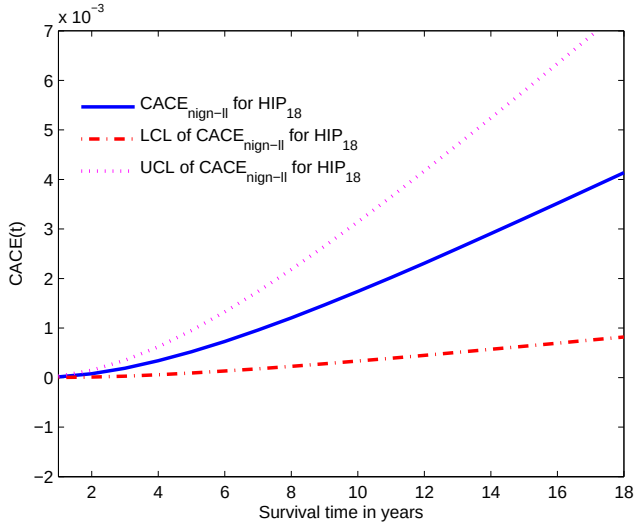

Figure 6.2: Estimated CACE(t) curve using the NIGN-LL model with 95% confidence interval: HIP_{18} data. The dash-dot line and the dotted line indicate the Lower Confidence Limit(LCL) and the Upper Confidence Limit (UCL).

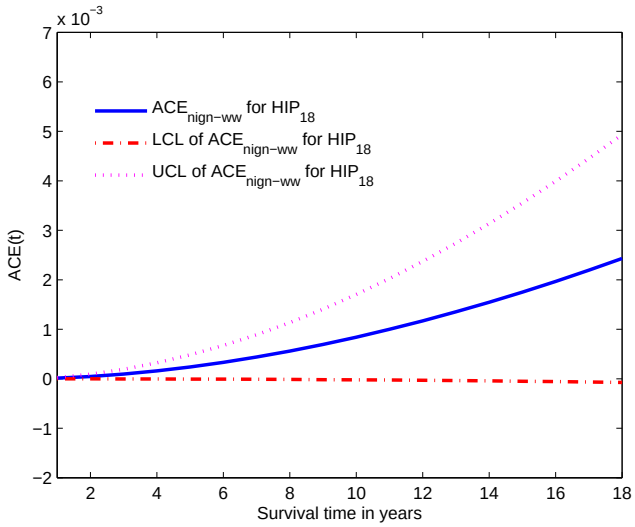

Figure 6.3: Estimated ACE(t) curve using the NIGN-WW model with 95% confidence interval: HIP_{18} data. The dash-dot line and the dotted line indicate the Lower Confidence Limit(LCL) and the Upper Confidence Limit (UCL).

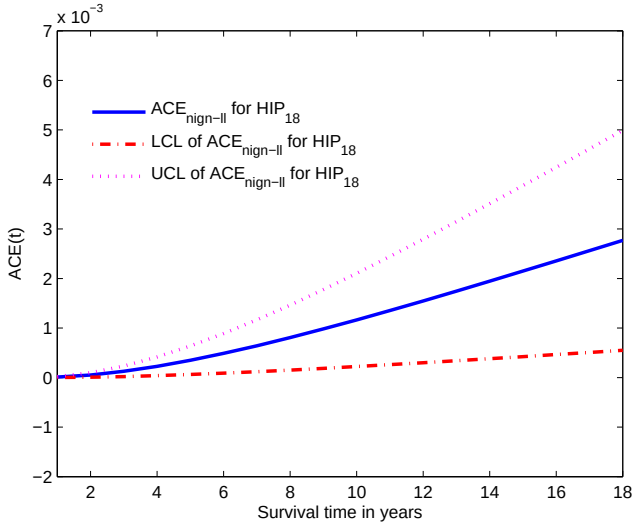

Figure 6.4: Estimated ACE(t) curve using the NIGN-LL model with 95% confidence interval: HIP_{18} data. The dash-dot line and the dotted line indicate the Lower Confidence Limit(LCL) and the Upper Confidence Limit (UCL).

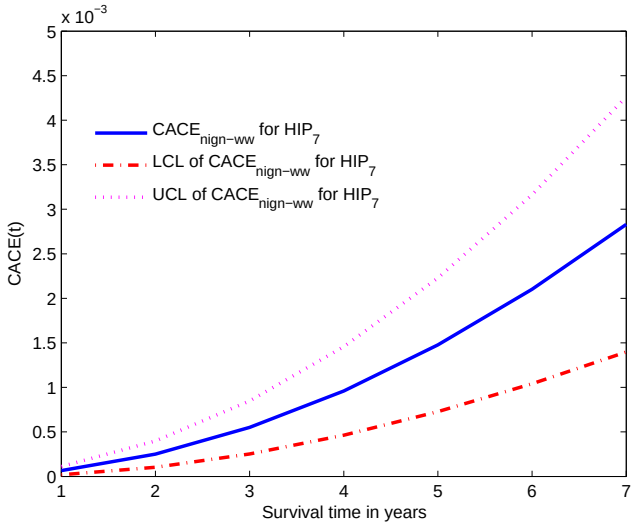

Figure 6.5: Estimated CACE(t) curve with 95% confidence interval: HIP_7 data using the NIGN-WW model. The dash-dot line and the dotted line indicate the Lower Confidence Limit(LCL) and the Upper Confidence Limit (UCL).

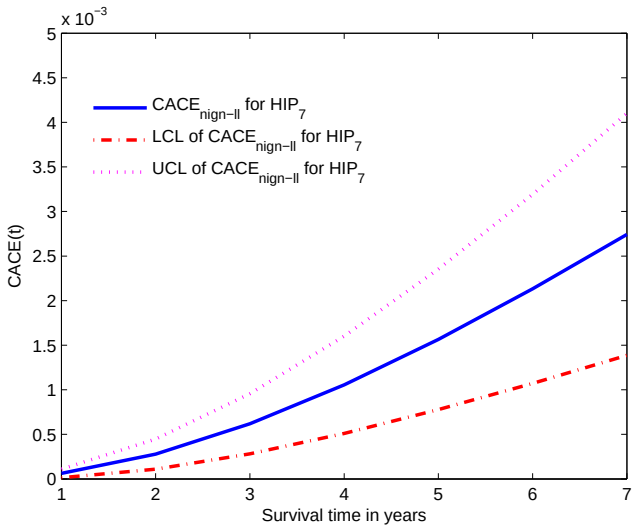

Figure 6.6: Estimated CACE(t) curve with 95% confidence interval: HIP_7 data using the NIGN-LL model. The dash-dot line and the dotted line indicate the Lower Confidence Limit(LCL) and the Upper Confidence Limit (UCL).

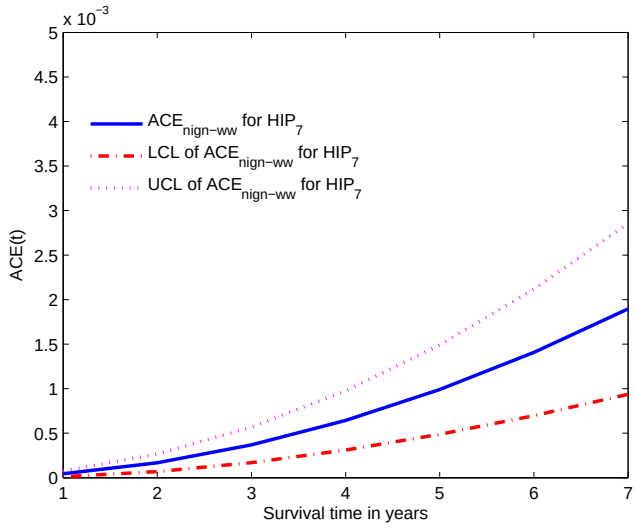

Figure 6.7: Estimated ACE(t) curve using the NIGN-WW model with 95% confidence interval: HIP_7 data. The dash-dot line and the dotted line indicate the Lower Confidence Limit(LCL) and the Upper Confidence Limit (UCL).

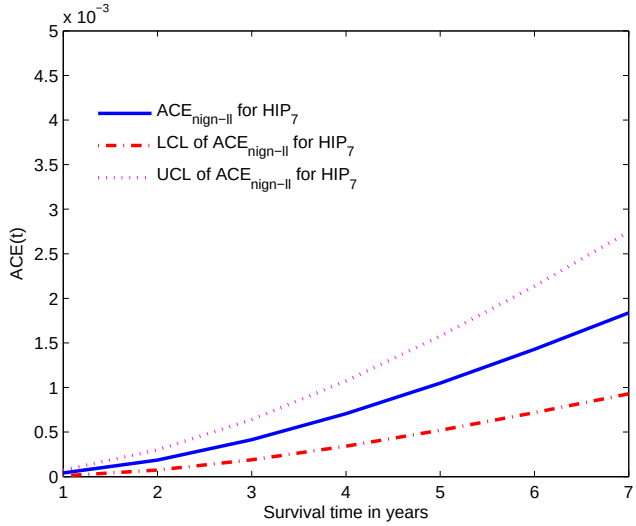

Figure 6.8: Estimated ACE(t) curve using the NIGN-LL model with 95% confidence interval: HIP_7 data. The dash-dot line and the dotted line indicate the Lower Confidence Limit(LCL) and the Upper Confidence Limit (UCL).

6.2.3 Estimation of relevant survival probabilities

The key point distinguishing our non-ignorable PPO-survival models (the NIGN-WW and NIGN-LL survival model) from conventional parametric survival models, is the fact that an individual's compliance type is allowed for, as proposed in Chapter 5. Unlike conventional parametric models, the NIGN-WW and the NIGN-LL models allow us to estimate the survival probability for three afore-mentioned subgroups in the HIP trial, namely Subgroup 1, Subgroup 2 and Subgroup 3 as defined in Section 6.2.1.

The estimated survival curves corresponding to these three subgroups (denoted by $S_{c1}(t)$, $S_{c0}(t)$ and $S_n(t)$) for the NIGN-WW and the NIGN-LL model application on the HIP_{18} and the HIP_7 subdata are given in Figure 6.9 (on page 148), where the top row gives the estimated survival curves evaluated using the NIGN-LL model for the HIP_{18} data (top left) and the HIP_7 data (top right). The figures located in the bottom row of Figure 6.9 represent the corresponding estimated curves for the NIGN-LL model. In addition, it is not difficult to obtain estimated survival probabilities from an ITT analytic perspective via our PPO-survival model; Figure 6.10 (on page 149) shows the relevant the ITT estimated survival curves. Equation (6.2.8) gives the formulations of calculating the ITT survival probabilities $S_0(t)$ and $S_1(t)$.

$$\begin{aligned} S_0(t) &= (1-\pi_c)S_n(t) + \pi_c S_{c0}(t); \\ S_1(t) &= (1-\pi_c)S_n(t) + \pi_c S_{c1}(t). \end{aligned} \tag{6.2.8}$$

All estimated values of the relevant survival probabilities, associated with Figure 6.9, and Figure 6.10 are detailed in Tables C-28, C-29, C-32 and C-33 in Appendix C for the reader's perusal.

The survival curves plotted on the two graphs at the right-hand side of Figure 6.9 (on page 148) correspond to the HIP_7 dataset via the NIGN-LL model (top graph) and the NIGN-WW model (bottom graph), respectively. They show that Subgroup 1, which consists of *compliers*, who are assigned to the screening (treated) group, has the highest survival probability among the three subgroups; Subgroup 2 has the lowest survival probability, in which individuals are all *compliers* and assigned to the control group; while the estimated survival curve of Subgroup 3, which consists of all *never-takers*,

who are assigned to either the screening or the control groups, lies between the curves of Subgroup 1 and Subgroup 2. These results can be interpreted thus: Subgroup 3 (*never-takers*) can be viewed as a healthy group, so it has a survival probability higher than Subgroup 2. Note that it is reasonable to make this assumption as presumably individuals in Subgroup 3 do not think a screening examination is necessary, for they refuse to take the screening examination (when assigned to the screening group). The estimated survival curves for the HIP_7 dataset also show that the NIGN-LL model and the NIGN-WW model provide very simliar results.

Unlike the analysis of the HIP_7 dataset, for the HIP_{18} dataset, the survival curves estimated using the NIGN-LL model and the NIGN-WW model are slightly different. The NIGN-WW model provides survival curves for Subgroups 2 and 3 that are closer than does the NIGN-LL model. However, the estimated survival curve for Subgroup 3 still lies between these two survival curves for Subgroup 1 and 2 (the left-hand graph in Figure 6.9). Therefore, the assumption mentioned above, that individuals in Subgroup 3 are all healthy — or at least thought of themselves as healthy at that time — is still valid. Since this is a scenario of long term follow-up, individuals' health status, at enrolment, may not affect their survival probability in later years. It is noteworthy that in the analysis of the HIP_{18} dataset, the survival curves of Subgroup 3 are closer to the survival curves of Subgroup 2 than Subgroup 1. This also makes sense in that individuals, either in Subgroup 2 or Subgroup 3, have not received a screening examination. It is thus not surprising to see these two subgroups having similar estimated survival curves. Indeed, the survival probability in later years is minimally related to the individual's health status at enrolment.

It is important to note that an ordinary ITT analysis does not provide these nuances nor this information (Figure 6.9, on page 148 and Figure 6.10 on page 149).

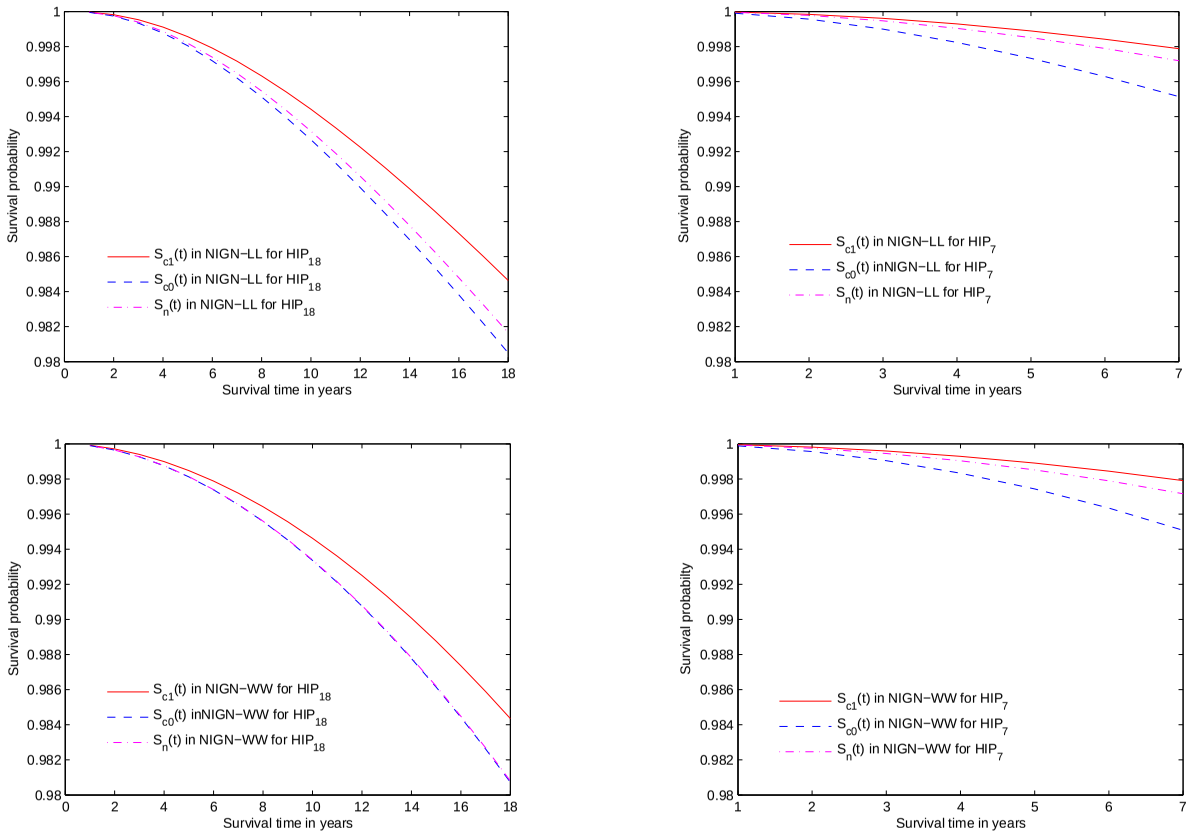

Figure 6.9: Estimated survival curves using the NIGN-LL model (top row) and the NIGN-WW model (bottom row) for the HIP_{18} (left) and HIP_7 (right) datasets. The solid line corresponds the survival curve for Subgroup 1: compliers assigned to the screening group; the dash line corresponds the survival curve for Subgroup 2: compliers assigned to the control group; the dash-dot line corresponds the survival curve for Subgroup 3: never-takers assigned to either the control or the screening groups.

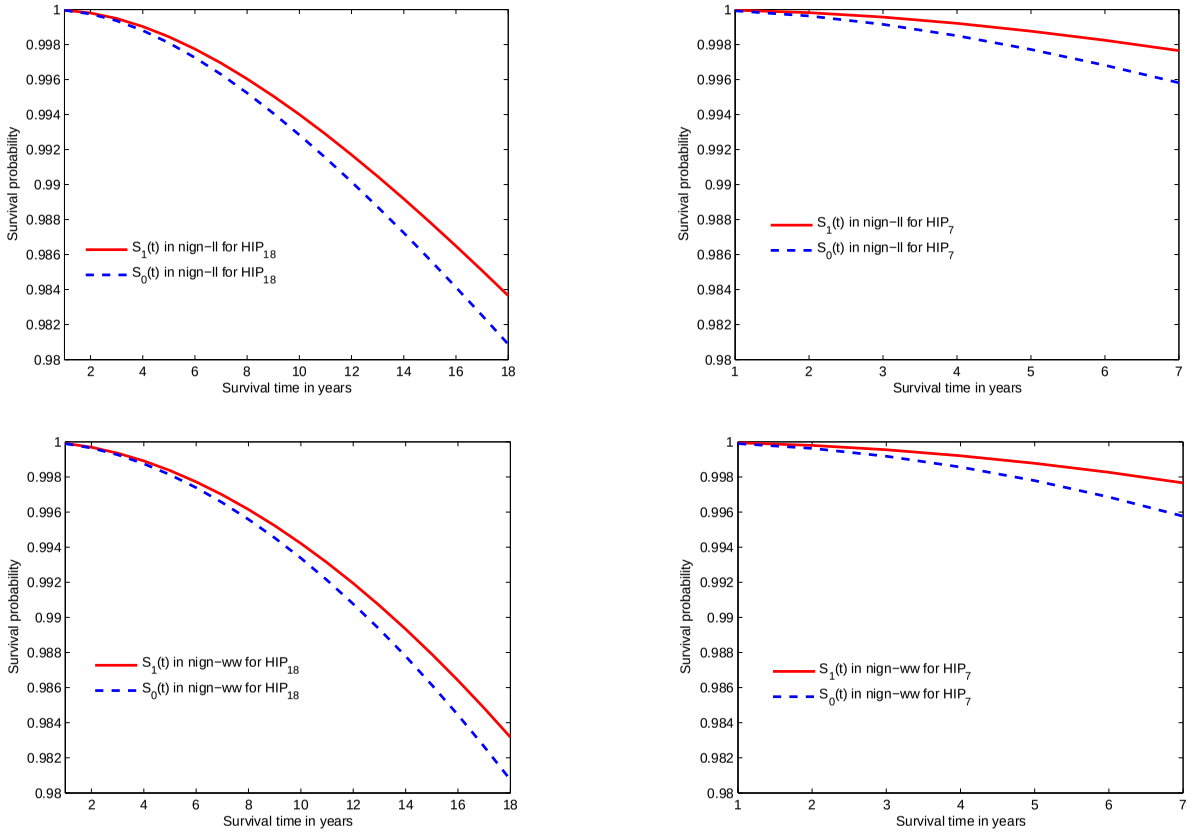

Figure 6.10: The ITT estimated survival curves using the NIGN-LL model (top row) and the NIGN-WW model (bottom row) for the HIP_{18} (left) and HIP_7 (right) datasets. The solid line corresponds the survival curve for the group assigned for screening ; the dash line corresponds the group not assigned for screening.

6.3 HIP data analysis: via the ignorable censoring mechanism PPO-survival models

In the IGN-W model, without pre-specifying a censoring distribution, the interest model parameter ($\boldsymbol{\theta}$) has the following form: $\boldsymbol{\theta}_W = (\rho_{c1},\ \rho_{c0},\ \rho_n,\ \alpha_T)^T$, where the first three elements, $\rho_{c1},\ \rho_{c0},\ \rho_n$, are the scale parameters of the Weibull distributions for failure time associated with the three subgroups (Subgroup 1, Subgroup 2 and Subgroup 3 given earlier); α_T is the common shape parameter for the three subgroups, and $\pi_{c,00}$ is the conditional probability of compliance given the assigned treatment ($Z_i = 0$) and the actual received treatment ($D_i(0) = 0$).

The HIP_7 data includes some individuals who were censored before year 7, the pre-set length of the follow-up period. This implies that the HIP_7 dataset does not satisfy the assumption of no censoring prior to the end of the follow-up. Since noncompliance occurred in the HIP trial, the compliance type has to be considered. Hence, we can not treat the HIP data as independent censoring. In other words, those assumptions required for ignorable PPO-survival model are invalid on the HIP_7 data. However, for the purpose of comparison, the IGN-W model is used to analyse the HIP_7 data, even though the latter assumption is violated. As in the non-ignorable PPO-survival model (detailed in Section 6.2), we firstly specify the parameter vector $\boldsymbol{\beta}$ for the IGN-W as is relevant to the HIP trial.

6.3.1 Model specification for the HIP trial

In the IGN survival model, the interest model parameter $\boldsymbol{\beta}$ can then be further simplified, as by definition no censoring distribution is involved in this model. That is:

$$\boldsymbol{\theta} = (\theta_{T,c1}, \theta_{T,c0}, \theta_{T,n}, \alpha_T, \pi_c), \tag{6.3.1}$$

The interest model parameter vector, $\boldsymbol{\theta}$, for the IGN survival model with a Weibull distribution (the IGN-W model) is specified in Table 6.5. This includes the three scale parameters, $\rho_{T,c1},\ \rho_{T,c0},\ \rho_{T,n}$, of the Weibull distribu-

Table 6.5: Ignorable PPO-survival model parameters for the HIP trial.

Model name	model parameter vector
IGN-WW	$(\rho_{T,c1},\ \rho_{T,c0},\ \rho_{T,n},\ \alpha_T)$

tions for the three subgroups and one common shape parameter, α, of the Weibull distributions.

The estimated parameter vector $\boldsymbol{\theta}$ of the IGN-W model for the HIP_{18} and the HIP_7 datasets are given in Tables C-6 and C-5 in Appendix C.

6.3.2 Estimation of causal effects: CACE(t) and ACE(t)

In this section, we evaluate the causal effect measured by CACE(t) and ACE(t) via the IGN-W model. The estimated $CACE(t)$ and $ACE(t)$ for the HIP_7 data (using the IGN-W model) are given in Tables 6.6 and 6.7 (on page 152). The first column indicates the survival time in years; the second column shows the estimated values of $CACE(t)$ or $ACE(t)$ at each observed year. The related standard errors are given in Column 3; the last two columns represent the lower and upper 95% confidence limits respectively. The related $CACE(t)$ curves are shown in Figure 6.11 (on page 153, for the HIP_{18} dataset) and in Figure 6.13 (on page 154, for the HIP_7 dataset). Note from Tables 6.6 and 6.7 that for all years ($t = 1, 2, \ldots, 7$) both the estimated CACE(t) and estimated ACE(t) derived from the IGN-W model are highly significant. Interpretation of these estimated CACE(t) and ACE(t) profiles is given later in Section 6.6.

Table 6.6: Estimated CACE(t) and its 95% CI using the IGN-W model for the HIP_7 data.

Method:	IGN-W		Data:	HIP_7		
Year	$\widehat{CACE}_{IGN-W}$	SE_{cace}	LCL_{cace}	UCL_{cace}	$Z-value$	$p-value$
1	0.82E−04	0.22E−04	0.40E−04	1.24E−04	3.7857	0.0001
2	3.13E−04	0.70E−04	1.76E−04	4.49E−04	4.4934	0.0000
3	6.85E−04	1.40E−04	4.10E−04	9.59E−04	4.8851	0.0000
4	11.9E−04	2.34E−04	7.35E−04	16.5E−04	5.1043	0.0000
5	18.4E−04	3.52E−04	11.5E−04	25.2E−04	5.2179	0.0000
6	26.1E−04	4.95E−04	16.4E−04	35.8E−04	5.2651	0.0000
7	35.1E−04	6.66E−04	22.0E−04	48.1E−04	5.2701	0.0000

HIP_7: HIP data with follow-up of 7 years

Table 6.7: Estimated ACE(t) and its 95% CI using the IGN-W model for the HIP_7 data.

Method:	IGN-W		Data:	HIP_7		
Year	$\widehat{ACE}_{IGN-W}$	SE_{ace}	LCL_{ace}	UCL_{ace}	$Z-value$	$p-value$
1	0.55E−04	0.15E−04	0.27E−04	0.83E−04	3.7852	0.0001
2	2.10E−04	0.47E−04	1.18E−04	3.01E−04	4.4927	0.0000
3	4.59E−04	0.94E−04	2.75E−04	6.43E−04	4.8842	0.0000
4	7.99E−04	1.57E−04	4.92E−04	11.1E−04	5.1032	0.0000
5	12.3E−04	2.36E−04	7.67E−04	16.9E−04	5.2167	0.0000
6	17.5E−04	3.32E−04	11.0E−04	24.0E−04	5.2639	0.0000
7	23.5E−04	4.46E−04	14.8E−04	32.2E−04	5.2688	0.0000

HIP_7: HIP data with follow-up of 7 years

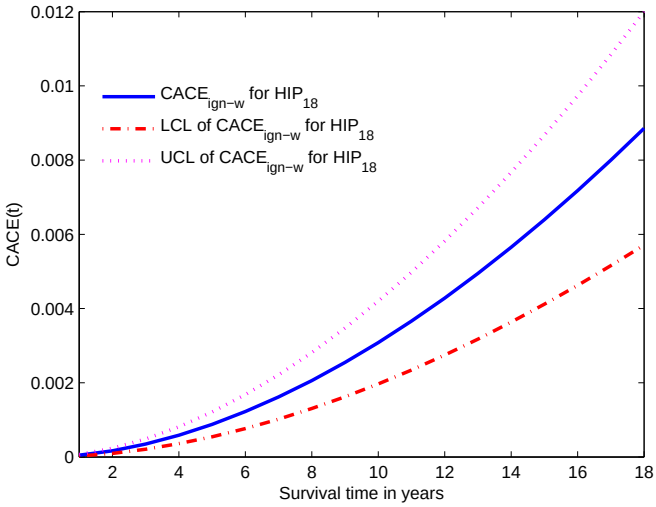

Figure 6.11: Estimated CACE(t) curve with 95% confidence interval: HIP_{18} data using the IGN-W model. The dash-dot line and the dotted line indicate the Lower Confidence Limit(LCL) and the Upper Confidence Limit (UCL).

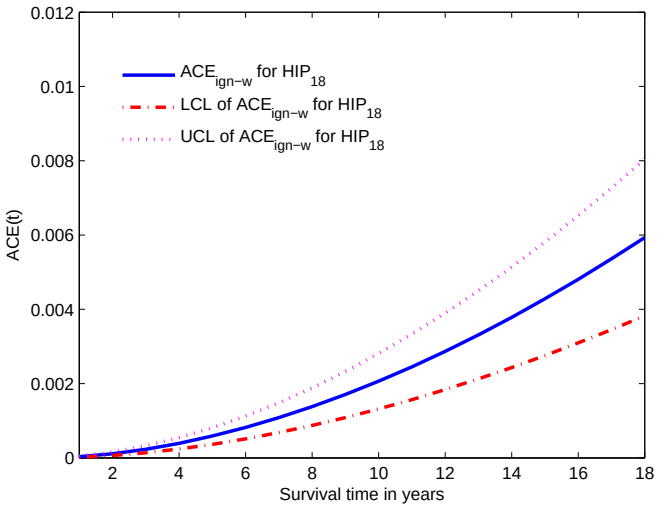

Figure 6.12: Estimated ACE(t) curve with 95% confidence interval: HIP_{18} data using the IGN-W model. The dash-dot line and the dotted line indicate the Lower Confidence Limit(LCL) and the Upper Confidence Limit (UCL).

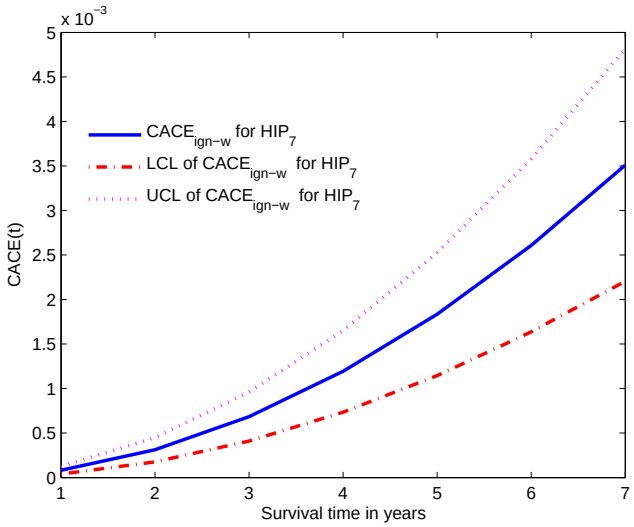

Figure 6.13: Estimated CACE(t) curve with 95% confidence interval: HIP_7 data using the IGN-W model. The dash-dot line and the dotted line indicate the Lower Confidence Limit(LCL) and the Upper Confidence Limit (UCL).

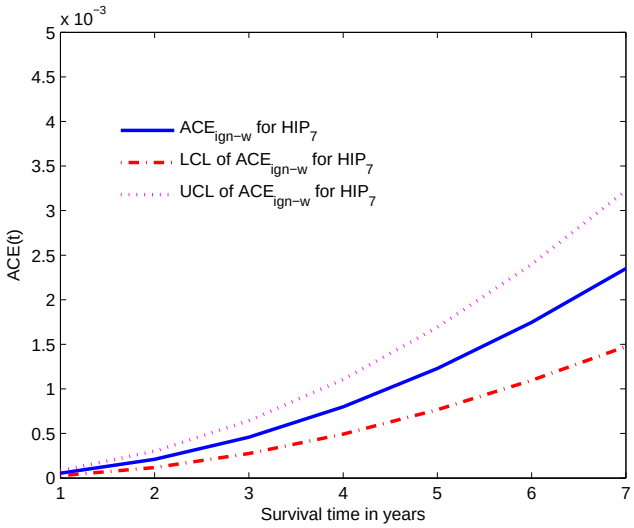

Figure 6.14: Estimated ACE(t) curve with 95% confidence interval: HIP_7 data using the IGN-W model. The dash-dot line and the dotted line indicate the Lower Confidence Limit(LCL) and the Upper Confidence Limit (UCL).

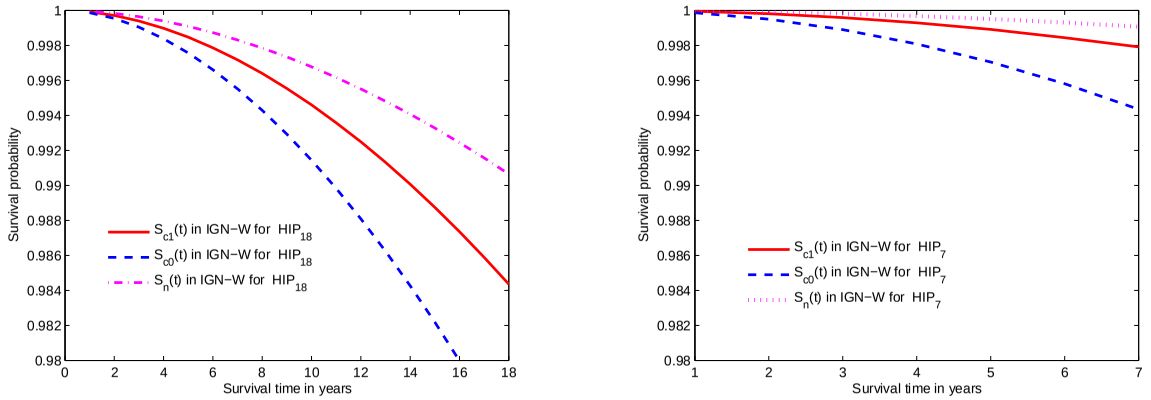

Figure 6.15: Estimated survival curves using the IGN-W model for the HIP_{18} data (left) and the HIP_7 data (right). The solid line corresponds to the survival curve for Subgroup 1: compliers assigned to the screening group; the dash line corresponds to the survival curve for Subgroup 2: compliers assigned to the control group; the dotted line corresponds to the survival curve for Subgroup 3: never-takers assigned to either the control or the screening groups.

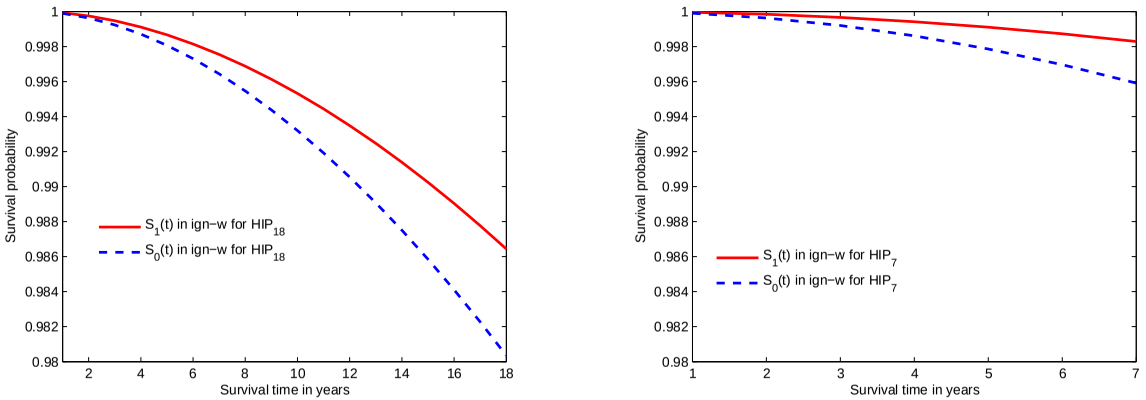

Figure 6.16: The ITT estimated survival curves using the IGN-W model for the HIP_{18} data (left) and the HIP_7 data (right). The solid line corresponds the survival curve for the group assigned for screening; the dash line corresponds the group not assigned for screening.

6.3.3 Estimation of relevant survival probabilities

Figure 6.15 (on page 155) gives the two estimated survival curves obtained via the IGN-W model. The left hand side indicates the estimated survival curves for the HIP_{18} data, and the right hand side shows that for the HIP_7 data. The exact numbers of these estimated survival curves (in Figure 6.15) are given in Tables C-37 and C-36 in Appendix C. Note that the estimated survival curves for the two datasets shown in Figure 6.15, correspond to three subgroups per dataset analysed. The three subgroups are: *compliers* from the assigned treatment group ($Z_i = 1$), *compliers* from the control group ($Z_i = 0$); and *never-takers* from either the assigned treatment group $Z_i = 1$ or the control group $Z_i = 0$. It is noteworthy that $\hat{S}_n(t)$ has a higher value than the other two subgroups, say $\hat{S}_{c1}(t)$ and $\hat{S}_{c0}(t)$. Recall that $\hat{S}_n(t)$ denotes the survival probability of individuals who are *never-takers* from either the assigned treatment group $Z_i = 1$ or the control group $Z_i = 0$. This information provided by Figure 6.15, $\hat{S}_n(t) > \hat{S}_{c1}(t) > \hat{S}_{c0}(t)$, seems counter-intuitive. It may occur because the HIP data does not satisfy the assumption required for the IGN-survival model, namely that no censoring occurs prior to end of the follow-up (Figure 6.15).

Figure 6.16 (on page 156) gives the relevant estimated survival curves for an ITT analysis using the IGN-W model for the HIP_{18} data (left) and the HIP_7 data (right); their associated estimated survival values are given in Tables C-39 and C-38 in Appendix C.

6.4 HIP data analysis: via the Frangakis-Rubin method

6.4.1 Frangakis-Rubin method

Let $S_{uz}(t)$ denote the survival function for the subgroup with $U_i = u$ and $Z_i = z$, and let $S_z(t)$ denote the survival function associated with two treatment levels, $z = 1$ for the treatment and 0 for the control. According to equation (5.1) given in Frangakis & Rubin (1999), the survival functions corresponding to the assigned treatment group $Z_i = 1$ (for $i \in \{i : Z_i = 1\}$) and the control group $Z_i = 0$ (for $i \in \{i : Z_i = 0\}$) can be re-written as follows:

$$\begin{aligned} S_0(t) &= (1 - \pi_c)S_{n0}(t) + \pi_c S_{c0}(t) \\ S_1(t) &= (1 - \pi_c)S_{n1}(t) + \pi_c S_{c1}(t); \end{aligned} \tag{6.4.1}$$

where π_c is the probability of compliers in the population.

Under the assumption of *exclusion restriction*, we have $S_{n0}(t) = S_{n1}(t)$. Using $S_n(t)$ to replace $S_{n0}(t)$ and $S_{n1}(t)$ in equation (6.4.1), we then have a form of the Frangakis-Rubin (F-R) model as follows:

$$\begin{aligned} S_0(t) &= (1 - \pi_c)S_n(t) + \pi_c S_{c0}(t); \\ S_1(t) &= (1 - \pi_c)S_n(t) + \pi_c S_{c1}(t). \end{aligned} \tag{6.4.2}$$

Note that $S_{c1}(t)$ and $S_n(t)$ can be estimated with the Kalan-Meier estimator directly.

Because compliance type $U_i = u$ is unobservable, $S_{c0}(t)$ the survival function associated with the subgroup with $U_i = c$ and $Z_i = 0$, cannot be estimated directly with the Kalan-Meier estimator through observation. Frangakis & Rubin (1999) provide a formulation for evaluating this, which is given below:

$$\hat{S}_{c0}(t) = \exp\left\{ - \int_0^t \frac{dN_0^{obs}(x)/N_0 - dN_{n1}(x)/N_1}{M_0(x)/N_0 - M_{n1}(x)/N_1} \right\}; \tag{6.4.3}$$

where N_1 and N_0 are the number of individuals randomly assigned to the treated and the control group respectively, and $N_i^{obs}(x)$ and $M_i^{obs}(x)$ denote

the stochastic processes (for $i = 1, \ldots, n$) and are defined as follows:

$$\begin{aligned} N_i(x) &:= I(T_i \leq x, R_i = 1); \\ M_i(x) &:= I(T_i \geq x); \end{aligned} \tag{6.4.4}$$

and

$$\begin{aligned} N_{n1}(x) &= \sum_i N_i(x) I(U_i = n) I(Z_i = 1); \\ N_0(x) &= \sum_i N_i(x) I(Z_i = 0); \\ M_{n1}(x) &= \sum_i M_i(x) I(U_i = n) I(Z_i = 1); \\ M_0(x) &= \sum_i M_i(x) I(Z_i = 0). \end{aligned} \tag{6.4.5}$$

Given equations (6.4.4) and (6.4.5), we note that $N_{n1}(x)$ and $M_{n1}(x)$ indicate the number of event occurrences or the number of survivors at or prior to the given time t for individuals who are never-takers assigned to the treatment group; whereas $N_0(x)$ and $M_0(x)$ denote the number of event occurrences or the number of survivors at the given time t for individuals who are assigned to the control group.

The Frangakis-Rubin (F-R) method (Frangakis & Rubin, 1999) can then be regarded as a non-parametric form of the noncompliance and non-ignorable survival potential outcomes model. This is true, because even though the F-R method does not, in fact, conventionally evaluate CACE(t) and ACE(t), the survival function values, $S_{c1}(t)$, $S_{c0}(t)$ and $S_n(t)$, which are required in the calculation of CACE(t) and ACE(t), can all be obtained by using the F-R model via the Kaplan-Meier estimator and via equations (6.4.3) and (6.4.2). More specifically, $S_{c1}(t)$ and $S_n(t)$), the survival probabilities corresponding to Subgroup 1 and Subgroup 3, can be estimated directly via the Kaplan-Meier estimator, and $S_{c0}(t)$, the survival probability corresponding to Subgroup 2, can be calculated using equation (6.4.3) as was proposed by Frangakis & Rubin (1999). Then $S_1(t)$ and $S_0(t)$, the survival probability corresponding to the assigned treatment $Z_i = 1$ and the control $Z_i = 0$ group, respectively, can be obtained directly via equation (6.4.2), which was also proposed by Frangakis & Rubin (1999).

The two datasets from the HIP trial, (HIP_{18} and HIP_7), are analysed according to the F-R method, using a programme written in MATLAB (Higham & Higham, 2000). This code follows equations (6.4.3) and (6.4.5) as proposed by Frangakis & Rubin (1999).

6.4.2 Estimation of relevant survival probabilities

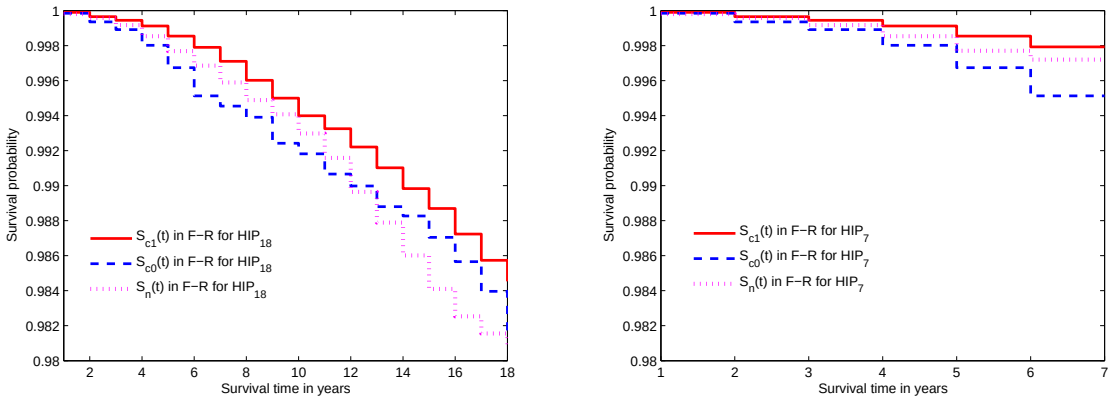

Figure 6.17: Estimated survival curves via the F-R for the HIP_{18} (left) and for the HIP_7 (right) data . The solid line corresponds to the survival curve for Subgroup 1: compliers assigned to the screening group; the dash-dot line corresponds to the survival curve for Subgroup 2: compliers assigned to the control group; the dotted line corresponds to the survival curve for the Subgroup 3: never-takers assigned to either the control or the screening group.

Figure 6.17 shows the estimated survival curves obtained via the F-R method, corresponding to the three subgroups ($S_{c1}(t)$, $S_{c0}(t)$ and $S_n(t)$) for the two HIP datasets, namely the HIP_{18} and HIP_7 subsets, where the left-hand graph corresponds to the HIP_{18} dataset, and the right-hand one to the HIP_7 dataset. Both of the graphs show that Subgroup 1 has a higher survival probability than the other two subgroups. This is similar to the results given in Section 6.2, which were obtained via the non-ignorable PPO-survival model. However, we can also see that in Figure 6.17, the estimated survival curves for Subgroup 3 are unlike those in Figure 6.9, especially for the HIP_{18}

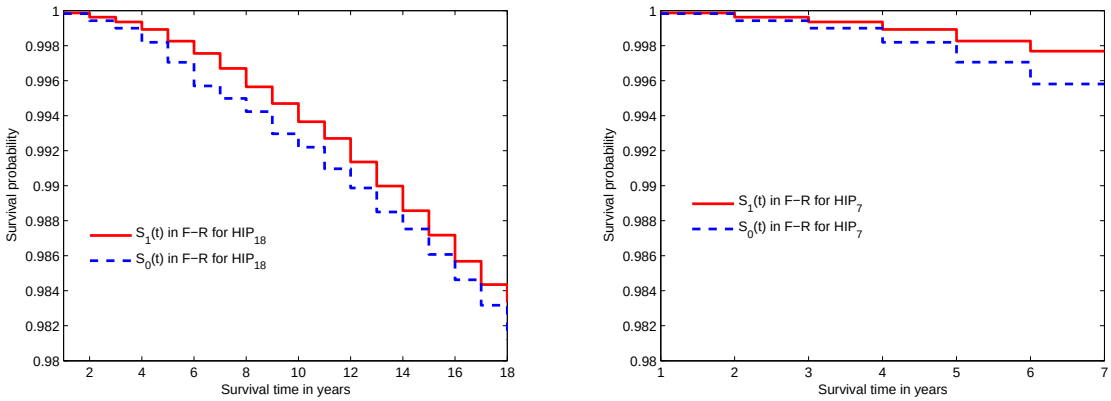

Figure 6.18: The ITT estimated survival curves via the F-R for the HIP_{18} (left) and for the HIP_7 (right) data . The solid line corresponds to the survival curve for the assigned screening group; the dash line corresponds to the survival curve for the assigned control group.

dataset. Figure 6.17 (left) (on page 160) shows that Subgroup 3 has a higher value than Subgroup 2 before 12 years, after which time ($t = 12$) Subgroup 3 has the lowest value among the three subgroups. These results are comprehensible, because the F-R model is a non-parametric based method, yet at the same time they are harder to interpret than results obtained via the non-ignorable PPO-survival model (Section 6.2.1).

While Figure 6.18 gives the estimated survival curves obtained using the F-R method in an ITT analysis, $S_1(t)$ and $S_0(t)$, correspond to the assigned treated group $Z_i = 1$ (the screen invited group) and the control group $Z_i = 0$, respectively. All estimated relevant survival values, which are associated with Figures 6.17 and 6.18, are given in Tables C-30, C-31, C-34 and C-35 in Appendix C.

6.4.3 Estimation of causal effects: CACE(t) and ACE(t)

The estimated values of CACE(t) (obtained by the F-R method) and the associated 95% confidence intervals (obtained by the bootstrap method (Efron & Tibshirani, 1993)) are given in Table 6.8 (for the HIP_7 dataset) and Ta-

ble 6.9 (for the HIP_{18} dataset) on page 163. Estimated ACE(t) values are given in Tables 6.10 and 6.11 (on page 164). Note that no parameters are estimated using the F-R model, as this is a non-parametric model. The formulae given at the end of Section 6.2.2, including equations (6.2.4), (6.2.6), (6.2.5) and (6.2.7), are also used to obtain Tables 6.8, 6.9, 6.10, 6.11, which provide estimated CACE(t) and ACE(t) values for the F-R method.

Related CACE(t) curves for the HIP_7 and the HIP_{18} data, estimated by the F-R method, are given in Figures 6.19 and 6.20 (on page 166).

In Tables 6.9 and 6.11, it is noteworthy that CACE(t) and ACE(t) for years 4 to 13 (inclusive), are significant in the F-R analyses of the HIP_{18}. This result is similar to that obtained for the HIP_7 data, where CACE(t) and ACE(t) for years 4 to 7 are highly significant (Tables 6.8 and 6.10).

Note that the estimated $CACE(t)$ at the first seven years for the HIP_7 and the HIP_{18} datasets (Table 6.9 and Table 6.8) are slight different. This is because of that these two datasets are not same even though we only consider the first seven years: 1) the censoring indicators for individuals whose observed time are seven years in these two HIP datasets are different. 2) the datasets re-sampled for Bootstrap are high likely not same. Recall that for the F-R method, the Bootstrap approach is employed for obtaining the standard errors of the estimated $CACE(t)$ and $ACE(t)$.

Table 6.8: Estimated CACE(t) (95% CI) using the F-R method for the HIP_7 data.

Method:	F-R		Data:	HIP_7		
Year	$\widehat{CACE}_{FR}$	SE_{cace}	LCL_{cace}	UCL_{cace}	Zvalue	$p-value$
1	0.57E−04	1.39E−04	−2.14E−04	3.29E−04	0.4127	0.3399
2	3.03E−04	2.43E−04	−1.74E−04	7.80E−04	1.2464	0.1063
3	5.35E−04	3.31E−04	−1.14E−04	11.8E−04	1.6159	0.0531
4	11.0E−04	4.44E−04	2.32E−04	19.7E−04	2.4836	0.0065
5	18.1E−04	5.85E−04	6.58E−04	29.5E−04	3.0846	0.0010
6	28.0E−04	7.08E−04	14.1E−04	41.9E−04	3.9505	0.0000
7	28.0E−04	7.08E−04	14.1E−04	41.9E−04	3.9505	0.0000

HIP_7: HIP data with follow-up of 7 years

Table 6.9: Estimated CACE(t) (95% CI) using the F-R method for the HIP_{18} data.

Method:	F-R		Data:	HIP_{18}		
Year	$\widehat{CACE}_{FR}$	SE_{cace}	LCL_{cace}	UCL_{cace}	Zvalue	$p-value$
1	0.57E−04	1.39E−04	−2.14E−04	3.29E−04	0.4127	0.3399
2	3.03E−04	2.43E−04	−1.74E−04	7.80E−04	1.2464	0.1063
3	5.35E−04	3.31E−04	−1.14E−04	11.8E−04	1.6159	0.0531
4	11.0E−04	4.44E−04	2.32E−04	19.7E−04	2.4836	0.0065
5	18.1E−04	5.85E−04	6.58E−04	29.5E−04	3.0846	0.0010
6	27.7E−04	7.13E−04	1.38E−03	41.7E−04	3.8930	0.0000
7	25.6E−04	7.91E−04	1.01E−03	41.1E−04	3.2321	0.0006
8	21.1E−04	8.68E−04	4.05E−04	38.1E−04	2.4272	0.0076
9	25.8E−04	9.69E−04	6.80E−04	44.8E−04	2.6616	0.0039
10	21.7E−04	10.5E−04	1.23E−04	42.3E−04	2.0772	0.0189
11	25.9E−04	11.3E−04	3.68E−04	48.0E−04	2.2849	0.0112
12	22.2E−04	12.2E−04	−1.64E−04	46.1E−04	1.8252	0.0340
13	22.1E−04	13.0E−04	−3.26E−04	47.5E−04	1.7082	0.0438
14	15.6E−04	13.8E−04	−11.3E−04	42.6E−04	1.1349	0.1282
15	16.5E−04	15.1E−04	−13.0E−04	46.1E−04	1.0962	0.1365
16	15.8E−04	16.3E−04	−16.1E−04	47.6E−04	0.9682	0.1665
17	17.7E−04	17.1E−04	−15.8E−04	51.1E−04	1.0353	0.1502
18	31.8E−04	18.2E−04	−3.96E−04	67.5E−04	1.7427	0.0407

HIP_{18}: HIP data with follow-up 18 years

Table 6.10: Estimated ACE(t) (95% CI) using the F-R method for the HIP_7 data.

Method:	F-R		Data:	HIP_7		
Year	$\widehat{ACE}_{FR}$	SE_{ace}	LCL_{ace}	UCL_{ace}	Zvalue	$p-value$
1	0.38E−04	0.93E−04	−1.44E−04	2.20E−04	0.4126	0.3399
2	2.03E−04	1.63E−04	−1.16E−04	5.23E−04	1.2464	0.1063
3	3.58E−04	2.22E−04	−0.76E−04	7.93E−04	1.6160	0.0530
4	7.38E−04	2.97E−04	1.56E−04	13.2E−04	2.4843	0.0065
5	12.1E−04	3.92E−04	4.41E−04	19.8E−04	3.0855	0.0010
6	18.7E−04	4.74E−04	9.44E−04	28.0E−04	3.9507	0.0000
7	18.7E−04	4.74E−04	9.44E−04	28.0E−04	3.9507	0.0000

HIP_7: HIP data with follow-up of 7 years

Table 6.11: Estimated ACE(t)(95% CI) using the F-R method for the HIP_{18} data.

Method:	F-R		Data:	HIP_{18}		
Year	$\widehat{ACE}_{FR}$	SE_{ace}	LCL_{ace}	UCL_{ace}	Zvalue	$p-value$
1	0.38E−04	0.93E−04	−1.44E−04	2.20E−04	0.4126	0.3399
2	2.03E−04	1.63E−04	−1.16E−04	5.23E−04	1.2464	0.1063
3	3.58E−04	2.22E−04	−0.76E−04	7.93E−04	1.6160	0.0530
4	7.38E−04	2.97E−04	1.56E−04	1.32E−03	2.4843	0.0065
5	12.1E−04	3.92E−04	4.41E−04	1.98E−03	3.0855	0.0010
6	18.6E−04	4.77E−04	9.23E−04	2.79E−03	3.8932	0.0000
7	17.1E−04	5.30E−04	6.74E−04	2.75E−03	3.2317	0.0006
8	14.1E−04	5.81E−04	2.71E−04	2.55E−03	2.4268	0.0076
9	17.3E−04	6.49E−04	4.55E−04	3.00E−03	2.6614	0.0039
10	14.6E−04	7.01E−04	0.82E−04	2.83E−03	2.0769	0.0189
11	17.3E−04	7.58E−04	2.46E−04	3.22E−03	2.2844	0.0112
12	14.9E−04	8.15E−04	−1.10E−04	3.08E−03	1.8249	0.0340
13	14.8E−04	8.68E−04	−2.19E−04	3.18E−03	1.7081	0.0438
14	10.5E−04	9.21E−04	−7.60E−04	2.85E−03	1.1350	0.1282
15	11.1E−04	10.1E−04	−8.72E−04	3.09E−03	1.0963	0.1365
16	10.5E−04	10.9E−04	−10.8E−04	3.19E−03	0.9683	0.1665
17	11.8E−04	11.4E−04	−10.6E−04	3.42E−03	1.0354	0.1502
18	21.3E−04	12.2E−04	−2.66E−04	4.52E−03	1.7426	0.0407

HIP_{18}: HIP data with follow-up of 18 years

6.5 Model comparisons

In Sections 6.2, 6.3 and 6.4, we analysed the HIP data using three models, namely the NIGN-WW model, the IGN-W model and the F-R method. In this section, we focus on comparing these three models using the results of analysis of both the HIP_{18} and HIP_7 datasets.

We first compare the non-ignorable PPO-survival model and the F-R method (Section 6.5.1). Secondly, we compare the ignorable PPO-survival model and the non-ignorable PPO-survival model (Section 6.5.2).

6.5.1 Non-ignorable PPO-survival model vs the F-R method

Since the non-ignorable PPO-survival model is developed under the same assumptions as those for the F-R method, it is reasonable to expect that the results from these two approaches would be comparable. However, it also would not be surprising to see some differences between these results on the HIP data; as indeed, these two approaches derive from different methodologies. Recall that the non-ignorable PPO-survival model is derived from a parametric perspective through the maximum likelihood method, whilst the F-R method is a nonparametric approach which uses the Kaplan-Meier estimator (p.84, Klein (1997)).

The results from the NIGN-WW and the NIGN-LL model are comparable to those obtained by the Frangakis-Rubin (F-R) method, as applied to the HIP_7 data (Figure 6.21, on page 167). This was also observed after an analogous analysis of the HIP_{18} data (Figure 6.23, on page 169).

Since the F-R method derives from a nonparametric perspective, it would not be surprising to see, however, some fluctuation in the F-R specific plot of CACE(t) (shown on the left hand side of Figure 6.23 for the HIP_{18} dataset). Note, however, from the left hand side plots (Figure 6.21), that little fluctuation is evident in the F-R specific CACE(t) plot for the HIP_7 dataset. This is because indeed the period of follow-up is shortened from 18 to 7 years.

In addition to the results mentioned above, we have shown that the non-ignorable PPO-survival models provide more precise estimators than those obtained by the F-R method (Figures 6.21, and 6.23), evident by the discrepancy in width of the associated confidence intervals.

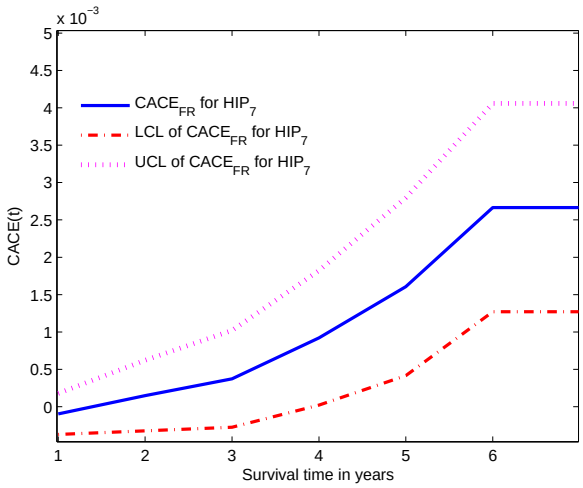

Figure 6.19: Estimated CACE(t) using the F-R method (with 95% confidence interval): HIP_7 data. The dash-dot line and the dotted line indicate the Lower Confidence Limit(LCL) and the Upper Confidence Limit (UCL).

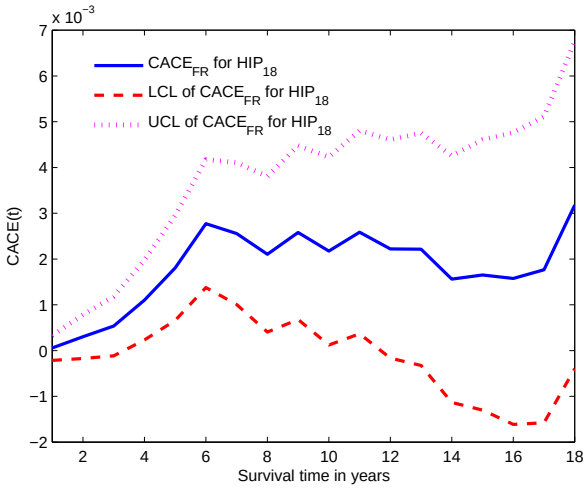

Figure 6.20: Estimated CACE(t) using the F-R method (with 95% confidence interval): HIP_{18} data. The dash-dot line and the dotted line indicate the Lower Confidence Limit (LCL) and the Upper Confidence Limit (UCL).

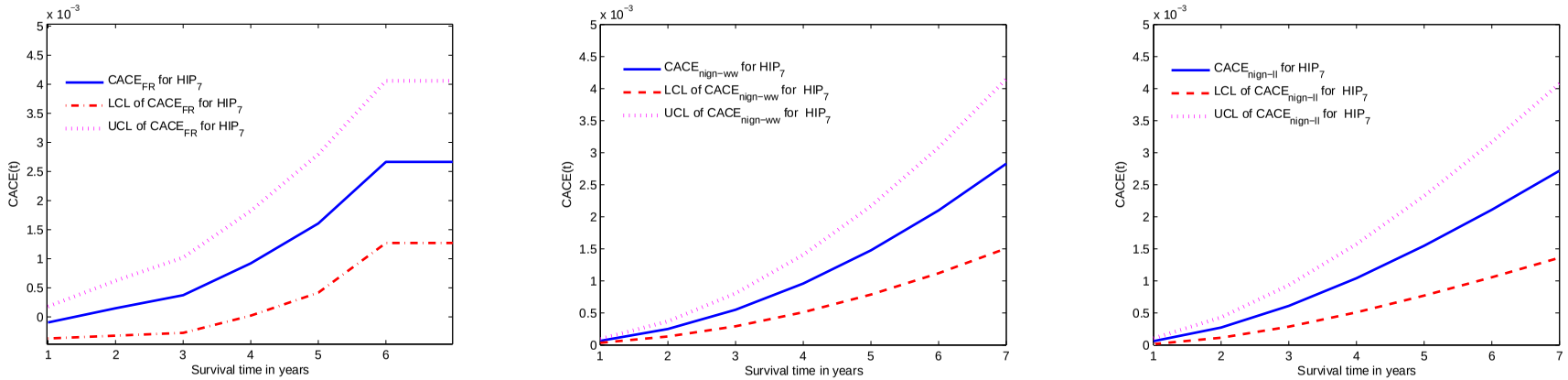

Figure 6.21: Estimated $CACE(t)$ curve with 95% confidence interval: HIP_7 data using the F-R method (left), the NIGN-WW model (middle) and the NIGN-LL model (right). The dash-dot line and dotted line indicate the Lower Confidence Limit(LCL) and the Upper Confidence Limit (UCL).

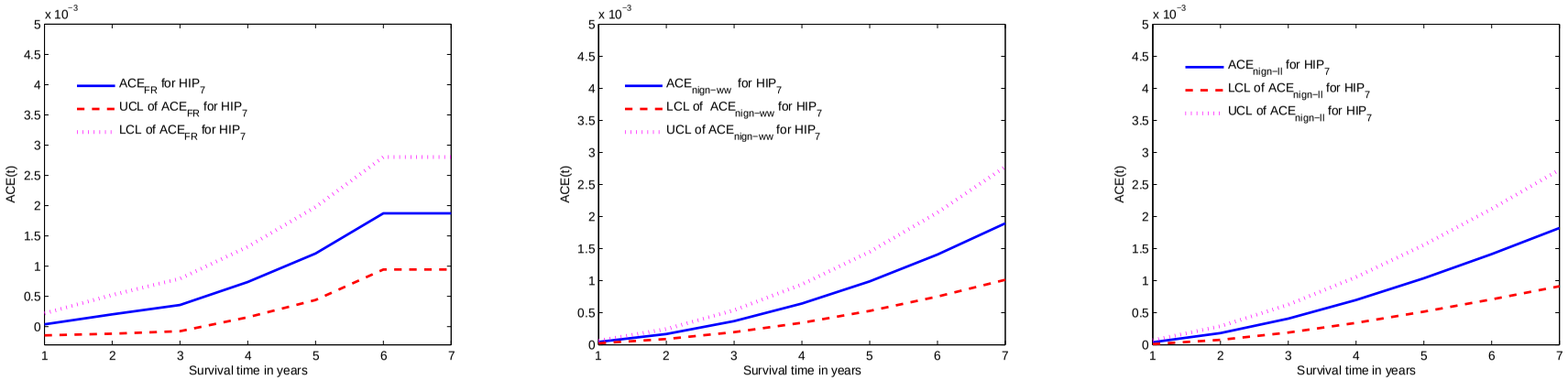

Figure 6.22: Estimated $ACE(t)$ curve with 95% confidence interval: HIP_7 data using the F-R method (left), the NIGN-WW model (middle) and the NIGN-LL model (right). The dash line and dotted line indicate the Lower Confidence Limit(LCL) and the Upper Confidence Limit (UCL).

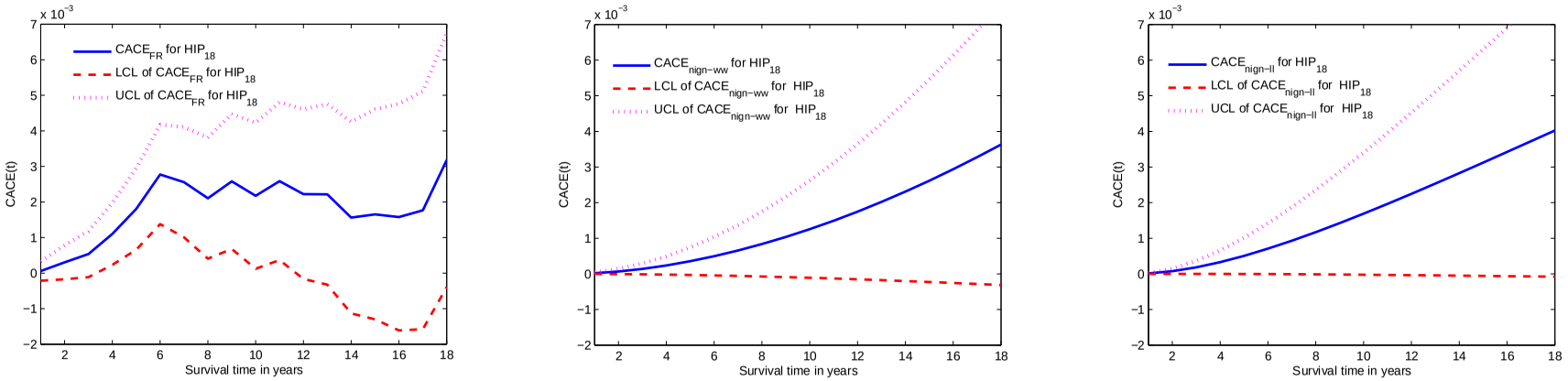

Figure 6.23: Estimated $CACE(t)$ curve with 95% confidence interval: HIP_{18} data using the F-R method (left), the NIGN-WW model (middle) and the NIGN-LL model (right). The dash-dot line and dotted line indicate the Lower Confidence Limit(LCL) and the Upper Confidence Limit (UCL).

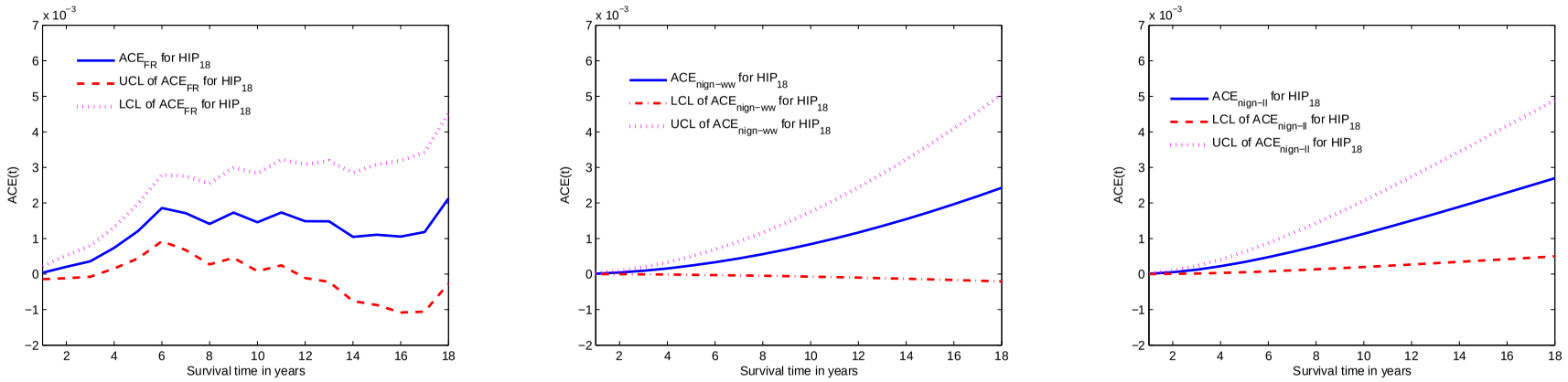

Figure 6.24: Estimated $ACE(t)$ curves with 95% confidence interval: HIP_{18} data using the F-R method (left), the NIGN-WW model (middle) and the NIGN-LL model (right). The dash line and dotted line indicate the Lower Confidence Limit(LCL) and the Upper Confidence Limit (UCL).

6.5.2 The NIGN-WW model vs the IGN-W model

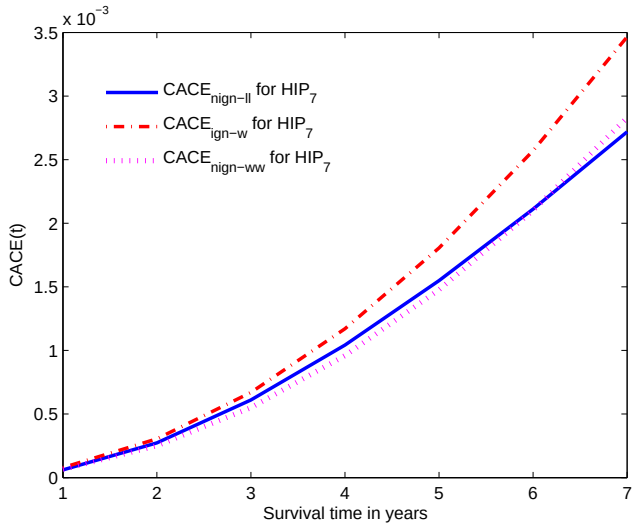

Figure 6.25: Comparisons of estimated CACE(t) curves using three models for the HIP_7 data. The dash-dot line shows the IGN-W model, the dotted line shows the NIGN-LL model and the solid line shows the NIGN-WW mdoel.

Figure 6.25 gives three model-specific curves of $\widehat{CACE}(t)$ for the HIP_7 data, including the $\widehat{CACE}(t)$ curve that estimated by the IGN-W model. It shows that the $\widehat{CACE}(t)$ curve corresponding to the IGN-W model is farther from the two curves associated with the NIGN-WW model and the NIGN-LL model. This implies that the IGN-W model may overestimate CACE(t) on the HIP_7 data. Hence, in the HIP_7 data, we need to assume latent ignorability; in which case, HIP_7 should be regarded as the special case of a non-ignorable censoring mechanism. Furthermore, Figure 6.25 shows that the two specific models from the non-ignorable PPO-survival model provide very similar values for $\widehat{CACE}(t)$ at each given time (in years). This provides strong evidence that the non-ignorable-survival model is a robust model, in that it is not sensitive to the assumed distribution. Column 2 of Table 6.12 gives the estimated CACE (for the HIP_7 data) with respect to failure time

Table 6.12: Comparisons between CACE and CACE(t) in two models: HIP_7

Model	$CACE_{\bar{T}}$	$CACE(1)$	$CACE(5)$	$AIC = -2ln(L(\theta)) + 2m$
NIGN-WW	52.64	0.66E-04	14.8E-04	2.0949E+05
NIGN-LL	1312.8	0.62E-04	15.7E-04	2.8393E+05

T (in the general definition, page 26), which can be expressed as

$$CACE = E(T_i(1)|Z_i = 1, U_i = c) - E(T_i(0)|Z_i = 0, U_i = c).$$

The values of CACE obtained by using the NIGN-WW model and the NIGN-LL model were 52.64 vs 1312.8. They are clearly showing that the estimator CACE strongly relies on the assumed distribution; while Column 3 and 4 of Table 6.12 give the estimated CACE(t) (in the specific definition, page 52) for $t = 1$ and $t = 5$ on the HIP_7 dataset via the two aforementioned models, namely $0.66E - 045$ vs $0.62E - 04$ (for $t = 1$) and $14.8E - 04$ vs $15.7E - 04$ (for $t = 5$). These results provide evidence that inferences for the specific *complier average causal effect* CACE(t) (for time-to-event data), are not as sensitive to distributional assumptions as inference for CACE. Indeed, by comparing their AIC (Akaike Information criterion [1]) for the HIP_7 data, the Weibull distribution ($AIC_{nign-ww} = 2.0948E + 5$) fits better than the log-normal ($AIC_{nign-ll} = 2.8392E + 5$) distribution for the non-ignorable PPO-survival model. But both models give similar results (Figures 6.5 and 6.6 on page 144).

On the other hand, estimated survival values for the HIP_{18} and HIP_7 data, using the IGN-W model (given on the right-hand graphs in Figure 6.26 on page 174 and Figure 6.28 on page 176), also show that the IGN-W model does not provide as reasonable results as the non-ignorable PPO-survival models, (the NIGN-WW and the NIGN-LL model). According to the estimated survival values for the three subgroups using the IGN-W model, Subgroup 3 is the highest among the three subgroups, which is not consistent

[1] A measure of the goodness fit of an estimated statistical model, $AIC = -2ln(L(\theta)) + 2m$ where m is the number of parameters and $ln(L(\theta))$ is the maximum nature logarithm likelihood function for the estimated model. The one with lowest AIC is the best model.

with the results provided via the F-R method and the non-ignorable PPO-survival models. This is hard to interpret reasonably.

In fact, both the NIGN-WW and the NIGN-LL models perform much better than the IGN-W model (Figure 6.28 on page 176) and Figure 6.26 on page 174).

Figures 6.26 (on page 174) and 6.28 (on page 176) provide strong evidence to support our argument, that in the case of both the HIP_{18} and HIP_{7} datasets, the NIGN-WW model provides more precise results than does the IGN-W model. With respect to the F-R and the NIGN-WW models, Subgroup 1 (individuals who took the screening examination) has a higher survival probability than subgroups where individuals did not take the screening examination. This shows a clear and significant treatment benefit of breast screening.

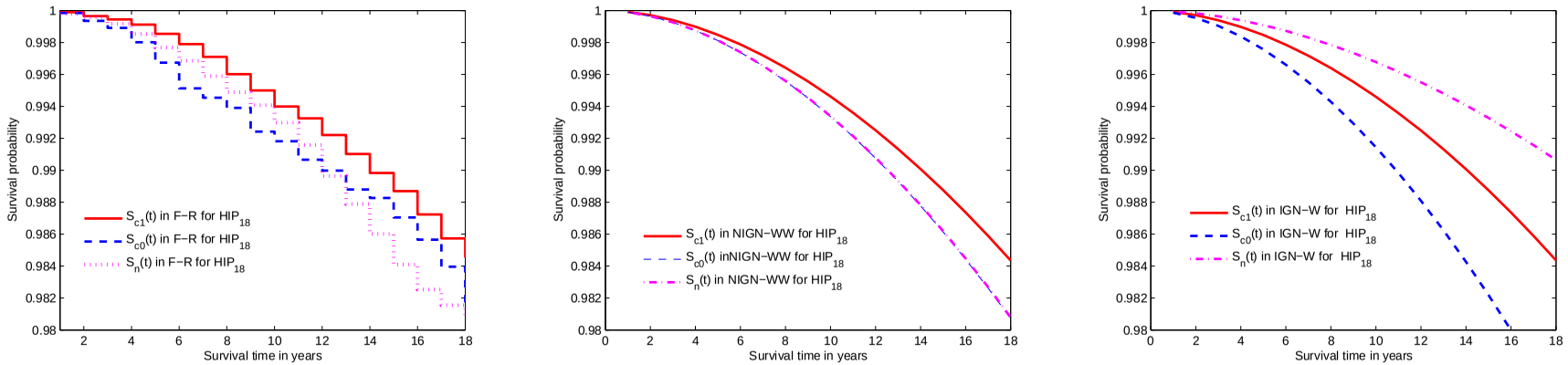

Figure 6.26: Estimated survival curves using the F-R (left), the NIGN-WW (middle) and the IGN-W (right) models for the HIP_{18} data. The solid line corresponds to the survival curve for Subgroup 1: compliers assigned to the screening group; the dash line corresponds to the survival curve for Subgroup 2: compliers assigned to the control group; the dash-dot line corresponds to the survival curve for the Subgroup 3: never-takers assigned to either the control or the screening group.

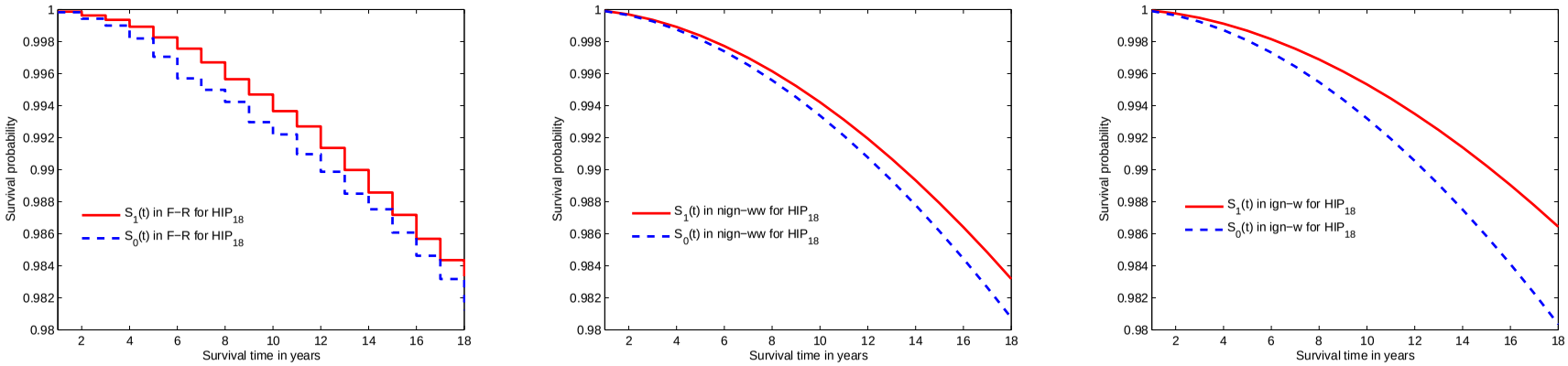

Figure 6.27: ITT estimated survival curves using the F-R (left), the NIGN-WW (middle) and the IGN-W (right) models for the HIP_{18} data. The solid line corresponds to the group assigned for screening; the dash line corresponds to the group not assigned for screening.

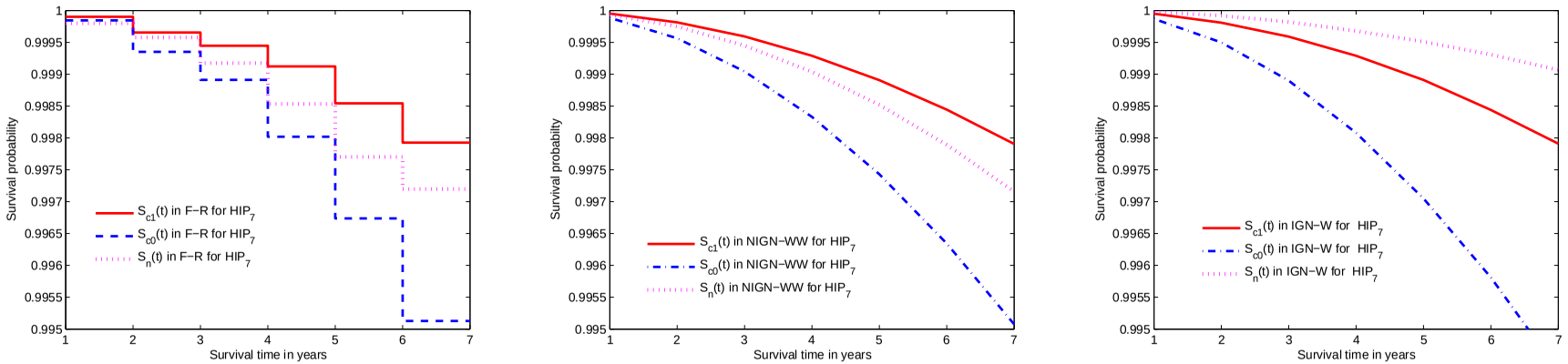

Figure 6.28: Estimated survival curves using the F-R (left), the NIGN-WW (middle) and the IGN-W (right) models for the HIP_7 data. The solid line corresponds to the survival curve for Subgroup 1: compliers assigned to the screening group; the dash-dot line corresponds to the survival curve for Subgroup 2: compliers assigned to the control group; the dotted line corresponds to the survival curve for Subgroup 3: never-takers assigned to either the control or the screening group.

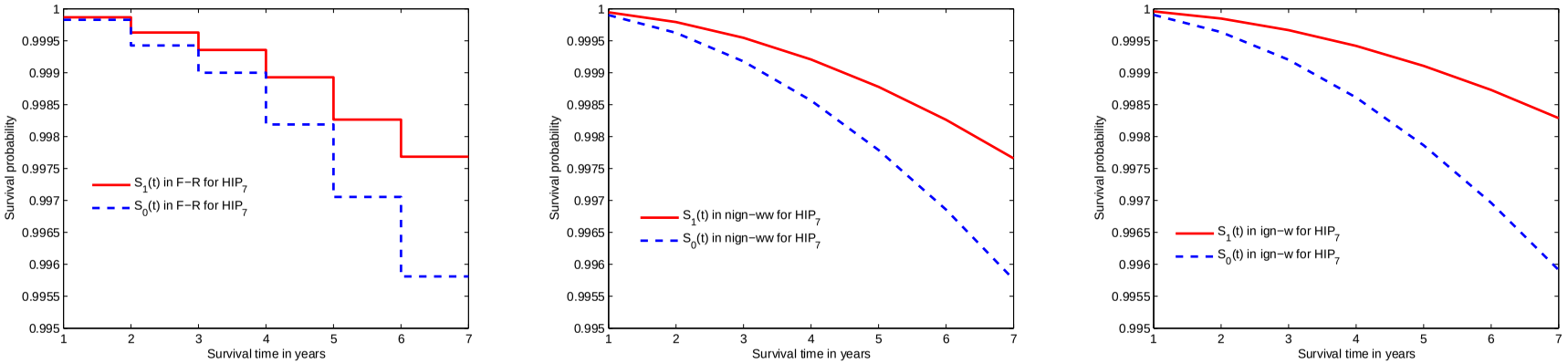

Figure 6.29: ITT estimated survival curves using the F-R (left), the NIGN-WW (middle) and the IGN-W (right) models for the HIP_7 data. The solid line corresponds to the group assigned for screening; the dash line corresponds to the group not assigned for screening.

6.6 Results of the analysis of the HIP data via the non-ignorable PPO-survival model

Table 6.3 (on page 139) shows the estimated complier average causal effect $\widehat{CACE}(t)$ and Table 6.4 (on page 140) shows the estimated average causal effect $\widehat{ACE}(t)$ obtained via the NIGN-WW and the NIGN-LL models. (Further details are given in Tables C-22, C-23, C-16 and C-17 in Appendix C). From these tables, we can see that $\widehat{CACE}(t)$ and $\widehat{ACE}(t)$ are completely positive at each given observed year. This implies that a woman undertaking screening would have a higher probability of surviving breast cancer than she would without having been screened.

6.6.1 Estimated CACE(t) for the HIP_7 data

More specifically, from the NIGN-WW model on the HIP_7 data, $\widehat{CACE}(5) = 14.8E-04$ and $SE_{cace(5)} = 3.83E-04$ (Table 6.3 on page 139 or Table C-22 on page 264). This indicates that after five years, a woman has about 0.148% more probability (with 0.0383% standard error) of survival than otherwise. This can also be interpreted as follows: out of 10,000, women about 14.8 ($\pm$3.83) women would be saved from breast cancer-related death five years after screening.

The result from the NIGN-LL model is similar, in that $\widehat{CACE}(5) = 15.7E-04$ ($SE_{cace(5)} = 4.02E-04$) (Table 6.3 on page 139 or Table C-23 on page 264). This is equivalent to about 15.7 ($\pm$4.02) women in 10,000 being saved from breast cancer-related death five years after screening.

6.6.2 Estimated CACE(t) for the HIP_{18} data

For the HIP_{18} data, the results show that in 10,000 women, about 3.56 ($\pm$1.86) women would be saved from breast cancer-related death five years after screening, as the complier average causal effect (CACE(t)) with its associated standard error estimated (via the NIGN-WW model) is $\widehat{CACE}(5) = 3.56E-04$ ($SE_{cace(5)} = 1.86E-04$) (Table 6.3 on page 139 or Table C-16 on page 260). On the other hand, the $\widehat{CACE}(t)$ obtained via the NIGN-LL model is $\widehat{CACE}(5) = 5.22E-04$ ($SE_{cace(5)} = 2.19E-04$) (Table 6.3 on page

139 or Table C-17 on page 261), which is equivalent to stating that about 5.22 ($\pm$2.19) women in 10,000 are being saved from breast cancer-related death five years after screening.

Even though the estimated values of $\widehat{CACE}(5)$ from these two specific non-ignorable PPO-survival models are different, they all tend to support the conclusion that women (over 40 years), who are screened would have a significantly reduced risk of breast cancer-related death.

By comparing the estimated complier average causal effects at the first year ($\widehat{CACE}(1)$) in the HIP trial, it is easy to see that $\widehat{CACE}(1) < \widehat{CACE}(5)$ (Table 6.3 on page 139) for the two specific non-ignorable PPO-survival models. More details are given in Tables C-22 , C-23, C-16 and C-17 in Appendix C. These results indicate that out of 10,000 women, the number of women saved from breast cancer related-death at the first year is significantly less than at the fifth year after screening. For the HIP_7 dataset, the number of women saved from breast cancer related-death at the first year is 0.66($\pm$0.24) vs 14.8($\pm$3.83) the number of women being saved at the fifth year (using the NIGN-WW model); with the NIGN-LL model, it is 0.62($\pm$0.25) vs 15.7($\pm$4.02).

This provides strong evidence to support, what is nowadays, accepted as common sense: the earlier breast cancer is detected, (and the earlier it is treated), the greater a woman's chance of survival. This conclusion is consistent with that reported by Shapiro et al. (1988). Mammography screening is now commonplace in the US (www.radiologyinfo.org), Australia (www.health.gov.au) and New Zealand (www.nzbcf.org.nz).

However, the results for the HIP_{18} and HIP_7 datasets (Table 6.3 on page 139) obtained using the NIGN-WW and the NIGN-LL models are different. One needs to make a decision as to which model is closer to the truth. As we stated in section 6.2.3, in a scenario of long-term follow-up, individuals's health status at beginning may not affect their survival probability in later years. Especially, Figure 6.17 (on page 160) shows that around 12 years it appears a crossing of the survival curves corresponding to Subgroup 2 $S_{c0}(t)$ and Subgroup 3 $S_n(t)$. This implies that for the HIP trial, the non-ignorable PPO-Survival models, (e.g. the NIGN-WW and the NIGN-LL models) would perform better in analysing a short term follow-up data (the HIP_7) than a

long-term follow-up data (the HIP_{18}).

Furthermore, by comparison with the F-R method, which gives $\widehat{CACE}(5) = 18.1E-04$ ($SE_{cace(5)} = 5.85E-04$) for both the HIP_{18} and HIP_7 datasets (Table 6.3 on page 139), we believe that the non-ignorable PPO-survival model is optimal for the HIP_7 data than the HIP_{18} data, because the results of the HIP_7 data obtained by the non-ignorable PPO-survival model, $\widehat{CACE}(5) = 14.8E-04$ ($SE_{cace(5)} = 3.83E-04$) (in the NIGN-WW model) and $\widehat{CACE}(5) = 15.7E-04$ ($SE_{cace(5)} = 4.02E-04$) (in the NIGN-LL model), are much more closer to those obtained via the F-R method than the results of the HIP_{18} data, which are $\widehat{CACE}(5) = 3.56E-04$ ($SE_{cace(5)} = 1.86E-04$) (in the NIGN-WW model) and $\widehat{CACE}(5) = 5.22E-04$ ($SE_{cace(5)} = 2.19E-04$) (in the NIGN-LL model).

Theoretically, the non-ignorable PPO-survival model and the F-R method are comparable, as they are both derived under the same assumptions. The difference between them is only that the former is a parametric survival model, while the latter is a non-parametric survival model. Table 6.13 (on page 180) shows the results of the HIP_7 data from the non-ignorable and F-R models. Clearly, the non-ignorable PPO-survival models provide significantly smaller standard errors for $\widehat{CACE}(t)$ compared to the F-R method.

Table 6.13: Results of the analysis of the HIP_7 data via the non-ignorable PPO survival model. Numbers indicate the number of increased survivals due to screening per 10,000 (on HIP_7 data). The associated standard error is given in brackets.

Survival year	NIGN-WW model	NIGN-LL model	F-R method
1	0.66($\pm$0.24)	0.62($\pm$0.24)	0.57 ($\pm$1.39)
5	14.8($\pm$3.83)	15.7($\pm$4.02)	18.1 ($\pm$5.85)

6.7 Summary

Briefly, the non-ignorable PPO-survival model variants do work well on the re-analysis of the HIP data, especially on the HIP_7 dataset using our two specific models, the NIGN-WW and the NIGN-LL model. This can be seen from the following: first, the results of CACE(t) and ACE(t) from these two models are comparable as shown by Tables C-22, C-23 (for CACE(t)) and by Tables C-20, C-21 (for ACE(t)). That is the NIGN-WW and the NIGN-LL specific CACE(t) and ACE(t) are very close in the analysis of the HIP_7 data (Figures 6.5 and 6.6 on page 144 for CACE(t) and Figures 6.7 and 6.8 on page 145 for ACE(t)). Secondly, these results are also comparable with the results obtained via the F-R method (Table 6.13 on page 180). Note, however, that the latter yields larger standard errors for its estimated CACE(t). In addition, inferences for ACE(t) and CACE(t) via the non-ignorable PPO-survival model are not sensitive to distributional assumptions for failure time and censoring time (Figure 6.25 on page 171 and Table 6.12 on page 172).

On the other hand, for RCTs with all-or-none compliance, the non-ignorable PPO-survival model provides more precise results of the analysis of the HIP_7 data than the ignorable PPO-survival model, because of the occurrence of noncompliance. Indeed, the ignorable PPO-survival model provides an overestimated treatment benefit (Figure 6.25 on page 171, Figure 6.15 on page 155, Table 6.6 and Table 6.7 on page 152). However, the ignorable PPO-survival model is worth consideration, as it can be applied to the scenario when the non-informative censoring assumption holds (Section 3.1.4).

The non-ignorable and ignorable PPO-survival models, proposed in this thesis, provide a significant contribution by dealing with the important issue of making causal inferences on survival outcomes from RCTs with all-or-none compliance. These models are derived by a likelihood-based method, which is different from the potential-outcome models proposed earlier by Robins and his colleagues (Robins & Tsiatis, 1991, 1992). Since our non-ignorable (and ignorable) PPO-survival models aim to analyse survival outcomes, we require different distributions rather than the conventional normal distribution. Thus our work differs from the approach of O'Malley & Normand (2005), which does not consider time-to-event data.

Our non-ignorable and ignorable PPO-survival models provide a new point of view in survival analysis, which combines the potential outcomes framework (Greenland, 2004; Pearl, 2001, 2000; Graham, 2000) with the traditional parametric survival approach (Klein, 1997; Cox & Oakes, 1984; Kalbfleisch & Prentice, 2002) and also adopts the EM algorithm (Dempster et al., 1977; Louis, 1982) to deal with unobservable variables. This is original work, not done to date. In Chapter 7, we extend the non-ignorable and ignorable PPO-survival models to the scenario when pre-treatment covariates are involved. Further discussion about the non-ignorable and ignorable PPO-survival models and their covariate extension are given in Chapter 8.

Chapter 7

Parametric potential-outcome survival models with pre-treatment covariates

7.1 Motivation

In Chapter 5, two parametric potential-outcome (PPO) survival models, namely the non-ignorable and ignorable censoring mechanism survival models, were formulated. They pertain, however, only to cases where no pre-treatment covariates (excluding assigned treatment Z_i) are included. In medical research, pre-treatment covariates, such as an individual's age, gender, smoking history, blood pressure, preliminary disease or health status etc., may play an important role in mediating the effect of treatment. For many RCTs, it is often of interest to investigate the causal effect of treatment, as the treatment effect may vary with pre-treatment covariate levels when these are involved (Peng et al., 2004; Mealli et al., 2004).

In this chapter, we extend the parametric potential-outcome (PPO) survival model(s) of Chapter 5 to incorporate time independent pre-treatment covariates, denoted by $\mathbf{X}$. Unlike Peng's approach (Peng et al., 2004), we model the compliance type variable (U_i) conditionally on related pre-treatment covariates ($\mathbf{X}$). This seems more natural than Peng's approach (Peng et al., 2004) which modelled the covariates, $\mathbf{X}$, conditionally on compliance type (Peng et al., 2004). However, the two different model specifications are, of course, related by Bayes' theorem. Note that to date, only Peng's work (Peng et al., 2004) has focused on estimating CACE and ACE in the presence of covariates, but Peng's approach is not applicable to time-to-event studies (Peng et al., 2004).

In this chapter, the assumptions specified in Chapter 5 are extended specifically to the case of time independent covariates, $\mathbf{X}$. Recall that, in

brief, these assumptions are: 1) the stable unit treatment value assumption (SUTVA), 2) random assignment, 3) exclusion restriction, 4) monotonicity, and 5) latent ignorablity. In addition, we assume that all pre-treatment covariates involved in a RCT may impact individual's compliance type. For convenience, we call the PPO-survival models derived in Chapter 5 the PPO-survival models (I), while the extensions derived in this chapter are called the PPO-survival model (II).

This chapter follows this outline: first we structure the joint distribution of all potential observable variables, including the pre-treatment covariates, in Section 7.2. The likelihood function of the PPO-survival model (II) is given in Section 7.3. Section 7.4 addresses the AFT model, and they are specified to Weibull and log-normal distributions for failure and censoring time in Section 7.5. The issues surrounding the model computation are given in Section 7.6. The special case of the extended PPO-survival model with pre-treatment covariates, the so-called ignorable PPO-survival model (II), is disccused in Section 7.7. Section 7.8 illustrates the implementation of the EM algorithm involved in the non-ignorable PPO-survival model (II) by using the HIP_7 data. Section 7.9 provides the summary of this chapter.

7.2 Structure of joint distribution with pre-treatment covariates

7.2.1 Joint distribution of $\mathbf{T}$*,* $\mathbf{C}$*,* $\mathbf{D}$*,* Z *and* X

In the scenario of additional pre-treatement covariates (X_i), for each individual, the potentially observable entities are given by $\underline{\mathbf{T}}$, $\underline{\mathbf{C}}$, $\underline{\mathbf{D}}$, $\underline{Z}$ and $\underline{X}$. As before, these variables are all assumed to be independent and identically distributed (i.i.d.), where the bold symbol with underline represent potential-outcome vectors for all individuals, such as $\underline{\mathbf{T}}$, $\underline{\mathbf{C}}$ and $\underline{\mathbf{D}}$. The joint distribution of the potentially observable entries can then be written as:

$$Pr(\underline{\mathbf{T}},\underline{\mathbf{C}},\underline{\mathbf{D}},\underline{Z},\underline{X};\boldsymbol{\Psi}_X) = \prod_{i=1}^{N} Pr(\mathbf{T}_i,\mathbf{C}_i,\mathbf{D}_i,Z_i,X_i;\boldsymbol{\Psi}_X), \tag{7.2.1}$$

where $\boldsymbol{\Psi}_X$ represents this new model's parameters, and $Pr()$ denotes the probability of variables given in brackets. By following the rule of conditional

probability, the joint distribution of all potential variables, say for the *ith* individual, can be written as below:

$$\begin{aligned} &Pr(\mathbf{T}_i,\mathbf{C}_i, \mathbf{D}_i, Z_i, X_i; \boldsymbol{\Psi}_X) \\ &\quad = Pr(\mathbf{T}_i, \mathbf{C}_i | Z_i, \mathbf{D}_i, X_i; \boldsymbol{\theta}_X) Pr(\mathbf{D}_i | Z_i, X_i; \varepsilon_X) Pr(Z_i; \nu) Pr(X_i; \omega). \end{aligned} \tag{7.2.2}$$

Note that $\boldsymbol{\Psi}_X$ is now broken down as $(\boldsymbol{\theta}_X, \varepsilon_X, \nu, \omega)$ in equation (7.2.2), where $\boldsymbol{\theta}_X$ denotes the parameter associated with the joint distribution of failure and censoring times that is conditioal on pre-treatment X_i ; ε_X represents the parameter associated with the conditional distribution of actual received treatment $\mathbf{D}_i$ (given the assigned treatment $Z_i = z$ and the pre-treatment covariates $X_i = \mathbf{x}$); and the parameters ν and ω are associated with the assigned treatment (Z_i) distribution and the pre-treatment covariates (X_i) distribution respectively.

Similar to the PPO-survival model (I) formulated in Chapter 5, the distribution of $\mathbf{D}_i = (D_i^{obs}, D_i^{mis})$, given the observed variables: assigned treatment Z_i and covariates X_i, can be re-written as follows:

$$\begin{aligned} Pr(\mathbf{D}_i | Z_i, \mathbf{X}_i; \varepsilon_{\mathbf{X}}) &= Pr(D_i^{obs}, D_i^{mis} | Z_i, X_i; \varepsilon_X) \\ &= Pr(D_i^{mis} | D_i^{obs}, Z_i, X_i; \varepsilon_{zdx}) Pr(D_i^{obs} | Z_i, X_i; \varepsilon_{zx}), \end{aligned} \tag{7.2.3}$$

where the parameter ε_X has been partitioned as $(\varepsilon_{zdx}, \varepsilon_{zx})$, for which the partitioned components denote the parameters corresponding to the conditional distributions of D_i^{mis} and D_i^{obs} separately. By the definition of actual received treatment D_i^{obs}, $Pr(D_i^{obs} | Z_i, X_i; \varepsilon_{zx})$ can be further split as follows:

$$Pr(D_i^{obs} | Z_i, X_i; \varepsilon_{zx}) = Pr(D_i^{obs} | Z_i, \varepsilon_z) Pr(Z_i; \nu) Pr(X_i; \omega). \tag{7.2.4}$$

The parameter ε_{zx} has then been broken down as $\varepsilon_{zx} = (\varepsilon_z, \nu, \omega)$. As stated in Section 5.3, for this extended model, we can re-write equation (5.3.2) to accommodate pre-treatment covariates as follows:

$$Pr(U_i | D_i^{obs}, Z_i, X_i; \boldsymbol{\pi}_X) \Leftarrow^{d} \Rightarrow Pr(D_i^{mis} | D_i^{obs}, Z_i, X_i; \varepsilon_{zdx}); \tag{7.2.5}$$

where $\Leftarrow^{d} \Rightarrow$ is used to denote distributional equivalence, as in equation

(5.3.2). Therefore, the joint distribution of all variables, including the potential-outcome vectors $\mathbf{T}_i$, $\mathbf{C}_i$, $\mathbf{D}_i$ and the observed variable Z_i and X_i, denoted by $Pr(\mathbf{T}_i, \mathbf{C}_i, \mathbf{D}_i, Z_i, X_i; \mathbf{\Psi}_X)$, is equivalent to the following product:

$$\begin{aligned} &Pr(\mathbf{T}_i,\mathbf{C}_i,\mathbf{D}_i,Z_i,X_i;\mathbf{\Psi}_X) \\ &\qquad \Leftarrow^{d} \Rightarrow \left\{ \begin{array}{l} Pr(\mathbf{T}_i,\mathbf{C}_i|Z_i,U_i,X_i;\boldsymbol{\theta}_X) \\ Pr(U_i|D_i^{obs},Z_i,X_i;\boldsymbol{\pi}_X) \\ Pr(D_i^{obs}|Z_i;\varepsilon_z)Pr(Z_i;\nu)Pr(X_i;\omega). \end{array} \right\} \end{aligned} \tag{7.2.6}$$

7.2.2 Joint distribution of the observable variables

The joint distribution of all observable variables, T_i^{obs}, C_i^{obs}, D_i^{obs}, Z_i and X_i, which is denoted by $Pr(T_i^{obs}, C_i^{obs}, D_i^{obs}, Z_i, X_i; \mathbf{\Psi}_X)$, can then be expressed as follows:

$$\begin{aligned} &Pr(T_i^{obs},C_i^{obs},D_i^{obs},Z_i,X_i;\mathbf{\Psi}_X) \\ &\quad = \int\int\int Pr(\mathbf{T}_i,\mathbf{C}_i,\mathbf{D}_i,Z_i,X_i;\mathbf{\Psi}_X)dD_i^{mis}dC_i^{mis}dT_i^{mis}. \end{aligned} \tag{7.2.7}$$

By substituting equation (7.2.6) into equation (7.2.7), the latter is equivalent to:

$$\begin{aligned} &Pr(T_i^{obs},C_i^{obs},D_i^{obs},Z_i,X_i;\mathbf{\Psi}_X) \\ &\Leftarrow^{d} \Rightarrow \int\int\int \left\{ \begin{array}{l} Pr(\mathbf{T}_i,\mathbf{C}_i|Z_i,U_i,X_i;\boldsymbol{\theta}_X) \\ \times Pr(U_i|D_i^{obs},Z_i,X_i;\pi_X) \\ \times Pr(D_i^{obs}|Z_i;\varepsilon_z)Pr(Z_i;\nu)Pr(X_i;\omega) \end{array} \right\} dU_i dC_i^{mis} dT_i^{mis} \end{aligned} \tag{7.2.8}$$

Note that T_i^{obs} and C_i^{obs} can never be fully observed together. So strictly, $Pr(T_i^{obs}, C_i^{obs}, D_i^{obs}, Z_i, X_i; \mathbf{\Psi}_X)$ cannot be considered to be the joint distribution of the observed variables. As in Section 5.4.2, we call it the joint distribution of the observable variable.

Furthermore, because the assumption of *latent ignorablility*, we have:

$$Pr(\mathbf{T}_i,\mathbf{C}_i|Z_i,U_i,X_i;\boldsymbol{\theta}_X) = Pr(\mathbf{T}_i|Z_i,U_i,X_i;\boldsymbol{\theta}_{TX})Pr(\mathbf{C}_i|Z_i,U_i,X_i;\boldsymbol{\theta}_{CX}). \tag{7.2.9}$$

Equation (7.2.8) can then be re-written as follows:

$$
\begin{aligned}
&Pr(T_i^{obs}, C_i^{obs}, D_i^{obs}, Z_i, X_i; \boldsymbol{\Psi}_X) \\
&\Leftarrow^{d}\Rightarrow \int\int\int \left\{ \begin{array}{l} Pr(\mathbf{T}_i|Z_i, U_i, X_i; \boldsymbol{\theta}_{TX}) \\ \times Pr(\mathbf{C}_i|Z_i, U_i, X_i; \boldsymbol{\theta}_{CX}) \\ \times Pr(U_i|D_i^{obs}, Z_i, X_i; \pi_X) \\ \times Pr(D_i^{obs}|Z_i; \varepsilon_z) Pr(Z_i; \nu) Pr(X_i; \omega) \end{array} \right\} dU_i dC_i^{mis} dT_i^{mis}
\end{aligned} \tag{7.2.10}
$$

Since compliance type U_i is a discreate variable with three catergories, equation (7.2.10) can be expressed as the similar form as equation (5.4.18), that is:

$$
\begin{aligned}
&Pr(T_i^{obs}, C_i^{obs}, D_i^{obs}, Z_i, X_i; \boldsymbol{\Psi}_X) \\
&\Leftarrow^{d}\Rightarrow \int\int\int \left\{ \begin{array}{l} Pr(\mathbf{T}_i, \mathbf{C}_i|Z_i, U_i, X_i; \boldsymbol{\theta}_X) \\ \times Pr(U_i|D_i^{obs}, Z_i, X_i; \boldsymbol{\pi}_X) \\ \times Pr(D_i^{obs}|Z_i; \varepsilon_z) Pr(Z_i; \nu) Pr(X_i; \omega) \end{array} \right\} dU_i dC_i^{mis} dT_i^{mis} \\
&= \left\{ \begin{array}{l} Pr(D_i^{obs}|Z_i; \varepsilon_z) \\ \times Pr(Z_i; \nu) \\ \times Pr(X_i; \omega) \end{array} \right\} \sum_{u\in\{a,c,n\}} \left\{ \begin{array}{l} \int Pr(\mathbf{T}_i|Z_i, U_i, X_i; \boldsymbol{\theta}_{TX}) dT_i^{mis} \\ \times \int Pr(\mathbf{C}_i|Z_i, U_i, X_i; \boldsymbol{\theta}_{CX}) dC_i^{mis} \\ \times Pr(U_i|D_i^{obs}, Z_i, X_i; \pi_X) \end{array} \right\} \\
&= Pr(D_i^{obs}, Z_i; \lambda) Pr(X_i; \omega) \sum_{u\in\{a,c,n\}} \left\{ \begin{array}{l} Pr(T_i^{obs}|Z_i, U_i, X_i; \boldsymbol{\theta}_{TX}) \\ \times Pr(C_i^{obs}|Z_i, U_i, X_i; \boldsymbol{\theta}_{CX}) \\ \times Pr(U_i|D_i^{obs}, Z_i, X_i; \pi_X) \end{array} \right\}
\end{aligned} \tag{7.2.11}
$$

7.3 The likelihood function

7.3.1 The idealised likelihood function

According to the formulation of the joint distribution of the observed variables, $Pr(T_i^{obs}, C_i^{obs}, D_i^{obs}, Z_i, X_i; \boldsymbol{\Psi}_X)$, given in equation (7.2.11), the idealised likelihood function for the PPO-survival model (II), denoted by $IDL_i^{(E)}(\boldsymbol{\Psi}_X)$,

can then be expressed as the following sums of products (in brackets):

$$
\begin{aligned}
&IDL_i^E(\boldsymbol{\Psi}_X) \\
&\quad = Pr(D_i^{obs}, Z_i, \lambda) Pr(X_i; \omega) \sum_{U_i \in \{c,n,a\}} \left\{ \begin{array}{l} Pr(T_i^{obs} | U_i, Z_i, X_i; \boldsymbol{\theta}_{TX}) \\ Pr(C_i^{obs} | U_i, Z_i, X_i; \boldsymbol{\theta}_{CX}) \\ Pr(U_i | D_i^{obs}, Z_i, X_i; \boldsymbol{\pi}_X), \end{array} \right\}
\end{aligned} \tag{7.3.1}
$$

where $\boldsymbol{\theta}_{TX}$ and $\boldsymbol{\theta}_{CX}$ denote the parameters associated with the failure and the censoring time distributions given the pre-treatment covariates X_i; $Pr(X_i; \omega)$ denotes the distribution of pre-treatment covariates X_i which is only included in the idealised likelihood function $IDL_i^E(\Psi_X)$ (equation (7.3.1)) rather than the idealised likelihood function $IDL_i(\Psi)$ (equation (5.5.2)); $Pr(D_i^{obs}, Z_i, \lambda)$ is same as in equation (5.5.2).

Note that except for $Pr(D_i^{obs}, Z_i, \lambda))$, the remains of distributions involved in the idealised likelihood function $IDL_i^E(\boldsymbol{\Psi}_X)$ (equation (7.3.1)), all are conditional on the given pre-treatment covariates X_i.

As in the PPO-survival model (I) (Chapter 5), we are interested in investigating the parameters associated with the failure time. We write the relevant parametrisation as $(\boldsymbol{\theta}_{TX})$ for this chapter's extension to covariate-adjusted PPO-survival models. However, from equation (7.3.1), it is clear that $\boldsymbol{\theta}_{TX}$ cannot be estimated independently from $\boldsymbol{\theta}_{CX}$ and $\boldsymbol{\pi}$, because these three paramters are all involved in the summation over compliance type and, therefore, the likelihood for these paramters does not separate into distinct components which can be maximised separately. We thus use the following notation $\boldsymbol{\beta} = (\boldsymbol{\theta}_{TX}, \boldsymbol{\theta}_{CX}, \boldsymbol{\pi}_X)$, to allow for all parametric inter-relationships. We focus now on deriving the idealised likelihood function with respect to $\boldsymbol{\beta}$ rather than with respect to $\boldsymbol{\Psi}_X$, denoted by $L_i^E(\boldsymbol{\beta})$. We then write $IDL_i^E(\boldsymbol{\beta})$ as follows:

$$
\begin{aligned}
&IDL_i^E(\boldsymbol{\beta}) \\
&\qquad = \sum_{U_i \in \{c,n,a\}} \left\{ \begin{array}{l} Pr(T_i^{obs} | U_i, Z_i, X_i; \boldsymbol{\theta}_{TX}) \\ \times Pr(C_i^{obs} | U_i, Z_i, X_i; \boldsymbol{\theta}_{CX}) \\ \times Pr(U_i | D_i^{obs}, Z_i, X_i; \boldsymbol{\pi}_X). \end{array} \right\}
\end{aligned} \tag{7.3.2}
$$

7.3.2 The conditional probability of compliance

Clearly, the term $Pr(U_i|D_i^{obs}, Z_i, X_i; \boldsymbol{\pi}_X)$ in equation (7.3.2), denotes the conditional probability of compliance type U_i, given $Z_i = z$, $D_i(z) = d$ and $X_i = \mathbf{x}$. Here, the parameter $\boldsymbol{\pi}_X$ can be specified as $\pi_{u,zdx} = Pr(U_i = u|Z_i = z, D_i(z) = d, X_i = \mathbf{x})$ for $u = (a, c, n),\ z,\ d = 0 \text{ or } 1, \text{ and } X_i = \mathbf{x}$.

Recall from Table 5.1 that knowledge of the values of $Z_i = z$ and $D_i = d$, allows the *ith* individual's compliance type to be determined for up to two types. We can thus assume that the conditional probability of compliance type, $U_i = u$, given $Z_i = z$, $D_i = d$, and $X_i = \mathbf{x}$, $\pi_{u,zdx}$, follows a logistic model (see Mealli et al. (2004) and, more recently, Mattei & Mealli (2007)). This can be expressed as follows:

$$logit(\pi_{c,zdx}) = \log(\frac{\pi_{c,zdx}}{1 - \pi_{c,zdx}}) = \eta_{zd0} + \eta_{zd1} X_i, \tag{7.3.3}$$

or equivalently:

$$\pi_{c,zdx} = \frac{\exp(\eta_{zd0} + \eta_{zd1} X_i)}{1 + \exp(\eta_{zd0} + \eta_{zd1} X_i)}. \tag{7.3.4}$$

where η_{zd0} and η_{zd1} denote the regression coefficients associated with the intercept term and the pre-treatment covariates X_i, respectively.

7.3.3 The likelihood function for complete data

Suppose that each individual's compliance type, $U_i = u$, is known. Then the idealised likelihood function, with respect to β, for the complete data is as follows:

$$IDL^E{}_i(\boldsymbol{\beta}, u) \quad = \left\{ \begin{array}{l} Pr(T_i^{obs}|U_i = u, Z_i = z, X_i = x; \boldsymbol{\theta}_{TX}) \\ \times Pr(C_i^{obs}|U_i = u, Z_i = z, X_i = x; \boldsymbol{\theta}_{CX}) \\ \times Pr(U_i = u|D_i^{obs} = d, Z_i = z, X_i = x; \pi_{u,zdx}). \end{array} \right\} \tag{7.3.5}$$

For the specified assigned treatment $Z_i = z$ and the pre-treatment covariates $X_i = \mathbf{x}$ (with known $U_i = u$), and given the indicator of censoring R_i, the

likelihood function of β, denoted by $L_i^E(\beta; u)$, has the following formulation:

$$
\begin{aligned}
&L^E{}_i(\boldsymbol{\beta}; u) = \\
&\left\{ \begin{array}{l}
\Pr(T_i(z) = t | U_i = u, Z_i = z, X_i = x; \boldsymbol{\theta}_{T,uzx})^{R_i} \\
\times \Pr(C_i(z) > t | U_i = u, , Z_i = z, X_i = x; \boldsymbol{\theta}_{C,uzx})^{R_i} \\
\times \Pr(T_i(z) > t | U_i = u, Z_i = z, X_i = x; \boldsymbol{\theta}_{T,uz})^{1-R_i} \\
\times \Pr(C_i(z) = t | U_i = u, , Z_i = z, X_i = x; \boldsymbol{\theta}_{C,uzx})^{1-R_i} \\
\times \Pr(U_i = u | D_i(z) = d, Z_i = z, X_i = x; \boldsymbol{\pi}_{u,zdx}).
\end{array} \right\}
\end{aligned}
\tag{7.3.6}
$$

Note that in equation (7.3.6), all relevant parameters have the pre-treatment covariates $\mathbf{x}$ included in their subscript. As in equation (5.5.6), R_i is the indicator of censoring, 1 indicates the event of interest, and 0 indicates censoring.

From equation (7.3.6), the *ith* individual with the pre-treatment $X_i = \mathbf{x}$ ($\mathbf{x} \in \mathcal{X}$) who is randomised to treatment $Z = 1$, and who dies or is censored at (observed) time t, contributes to the complete data likelihood for $\boldsymbol{\beta}$ as follows:

$$
\begin{aligned}
&L^E{}_{i \in \{i: Z_i = 1\}}(\boldsymbol{\beta}; u) \\
&\quad = \left\{ \begin{array}{l}
\Pr(T_i(1) = t | U_i = u, Z_i = 1, X_i = x; \boldsymbol{\theta}_{T,u1x})^{R_i} \\
\times \Pr(C_i(1) > t | U_i = u, Z_i = 1, X_i = x; \boldsymbol{\theta}_{C,u1x})^{R_i} \\
\times \Pr(T_i(1) > t | U_i = u, Z_i = 1, X_i = x; \boldsymbol{\theta}_{T,u1x})^{1-R_i} \\
\times \Pr(C_i(1) = t | U_i = u, Z_i = 1, X_i = x; \boldsymbol{\theta}_{C,u1x})^{1-R_i} \\
\times \Pr(U_i = u | D_i(1) = d, Z_i = 1, X_i = x, \boldsymbol{\pi}_{u,1dx});
\end{array} \right\}
\end{aligned}
\tag{7.3.7}
$$

for $d = 0$ or 1, $u \in \{c, a, n\}$ and $x \in \mathcal{X}$.

Similarly, the *ith* individual with the pre-treatment $X_i = \mathbf{x}$ ($x \in \mathcal{X}$), who is randomised to the control $Z = 0$, and who dies or is censored at (observed) time t, contributes to the complete data likelihood for $\boldsymbol{\beta}$ as follows:

$$
\begin{aligned}
&L^E{}_{i \in \{i: Z_i = 0\}}(\boldsymbol{\beta}; u) \\
&\quad = \left\{ \begin{array}{l}
\Pr(T_i(0) = t | U_i = u, Z_i = 0, X_i = x; \boldsymbol{\theta}_{T,u0x})^{R_i} \\
\times \Pr(C_i(0) > t | U_i = u, Z_i = 0, X_i = x; \boldsymbol{\theta}_{C,u0x})^{R_i} \\
\times \Pr(T_i(0) > t | U_i = u, Z_i = 0, X_i = x; \boldsymbol{\theta}_{T,u0x})^{1-R_i} \\
\times \Pr(C_i(0) = t | U_i = u, Z_i = 0, X_i = x; \boldsymbol{\theta}_{C,u0x})^{1-R_i} \\
\times \Pr(U_i = u | D_i(0) = d, Z_i = 0, X_i = x; \boldsymbol{\pi}_{u,0dx});
\end{array} \right\}
\end{aligned}
\tag{7.3.8}
$$

for $d = 0$ or 1, ,$u \in \{c,\ a,\ n\}$ and $x \in \mathcal{X}$.

Hence, given the values of z and d, equations (7.3.7) and (7.3.8) can then be expressed as one overall equation with the following formulation:

$$
\begin{aligned}
&L^{E}{}_{i\in\{i:Z_i=z,D_i(z)=d\}}(\boldsymbol{\beta};u) \\
&= \left\{ \begin{array}{l}
\Pr(T_i(z)=t|U_i=u,Z_i=z,X_i=x;\boldsymbol{\theta}_{T,uzx})^{R_i} \\
\times\Pr(C_i(z)>t|U_i=u,Z_i=z,X_i=x;\boldsymbol{\theta}_{C,uzx})^{R_i} \\
\times\Pr(T_i(z)>t|U_i=u,Z_i=z,X_i=x;\boldsymbol{\theta}_{T,uzx})^{1-R_i} \\
\times\Pr(C_i(z)=t|U_i=u,Z_i=z,X_i=x;\boldsymbol{\theta}_{C,uzx})^{1-R_i} \\
\times\Pr(U_i=u|D_i(z)=d,X_i=x,\pi_{u,zdx});
\end{array} \right\}
\end{aligned}
\tag{7.3.9}
$$

for z, $d = 0$ or 1, $u \in \{c,\ a,\ n\}$ and $x \in \mathcal{X}$.

As before, for notational convenience, we introduce the following conventions in this chapter:

$$
\begin{aligned}
f^i_{ux}(v,t;\theta_{T,uvx}) &= pr(T_i(v)=t|U_i=u,X_i;\theta_{T,uvx}), \quad \text{for } v=0,\ 1, \\
e^i_{ux}(v,t;\theta_{C,uvx}) &= pr(C_i(v)=t|U_i=u,X_i;\theta_{C,uvx}), \quad \text{for } v=0,\ 1, \\
S^i_{ux}(v,t;\theta_{T,uvx}) &= Pr(T_i(v)>t|U_i=u,X_i;\theta_{T,uvx}), \quad \text{for } v=0,\ 1, \\
G^i_{ux}(v,t;\theta_{C,uvx}) &= Pr(C_i(v)>t|U_i=u,X_i;\theta_{C,uvx}), \quad \text{for } v=0,\ 1,
\end{aligned}
$$

where $Pr()$ and $pr()$ represent the probability density distribution and the probability distriution, respectively, and v dentoes potential treatment levels.

Using this notation in equation (7.3.9), we obtain the form of the likelihood function for the *ith* individual with respect to $\boldsymbol{\beta}$ for the non-ignorable PPO-survival model (II), which is given in equation (7.3.10) below. This corresponds to the case where the pre-treatment covariates X_i and compliance type U_i are assumed to be known. We write this as:

$$
\begin{aligned}
&L^{E}{}_{i\in\{i:Z_i=z,D_i(z)=d\}}(\boldsymbol{\beta};u) \\
&= \left\{ \begin{array}{l}
[f^i_{ux}(z,t;\theta_{T,uzx})G^i_{uzx}(z,t;\theta_{C,uzx})]^{R_i} \\
[S^i_{ux}(z,t;\theta_{T,uzx})e^i_{ux}(v,t;\theta_{C,uzx})]^{1-R_i}\ \pi_{u,zdx}.
\end{array} \right\}
\end{aligned}
\tag{7.3.10}
$$

Equation (7.3.10) thus gives the likelihood function (corresponding to the *ith* individual) for complete data, in the so-called non-ignorable PPO-survival

model (II), where the model parameter $\boldsymbol{\beta}$ accommodates the inclusion of covariates X, which has a form as the following:

$$\boldsymbol{\beta} = (\boldsymbol{\theta}_{T,uzx}, \boldsymbol{\theta}_{C,uzx}, \boldsymbol{\pi}_{u,zdx}); \tag{7.3.11}$$

where $\boldsymbol{\theta}_{T,uzx}$ denotes the parameter associated with the failure time distribution, it can be further partitioned when considering the specific subgroups, that is

$$\boldsymbol{\theta}_{T,uzx} = (\boldsymbol{\theta}_{T,c1x}, \boldsymbol{\theta}_{T,c0x}, \boldsymbol{\theta}_{T,nx}, \boldsymbol{\theta}_{T,ax}). \tag{7.3.12}$$

Similarly, the parameter associated with the censoring time distributioin $\boldsymbol{\theta}_{C,uzx}$ can be expressed as follows:

$$\boldsymbol{\theta}_{C,uzx} = (\boldsymbol{\theta}_{C,c1x}, \boldsymbol{\theta}_{C,c0x}, \boldsymbol{\theta}_{C,nx}, \boldsymbol{\theta}_{C,ax}). \tag{7.3.13}$$

Note that, unlike in the non-ignorable PPO-survival model (I), the parameter associated with the conditional distribution of compliance type is denoted by $\pi_{u,zdx}$ rather than $\pi_{u,zd}$ because in this case the pretreatment covariates X are involved. This implies that more parameters have to be included in the model paremeter $\boldsymbol{\beta}$. In fact, equation (7.3.4) gives the formula of calculating $\pi_{c,zdx}$, where η_{zd0} and η_{zd1} (for both z and d with 1 and 0) refer to the regression coefficients in the logistic model.

7.4 AFT model including pre-treatment covariates

For time-to-event studies, the accelerated failure time model (AFT) (p.64 Cox & Oakes (1984) or p.46, Klein (1997)) is a powerful tool to link survival outcomes with related covariates. Recall that in Section 5.7, we derived an AFT model for a RCT with noncompliance without covariates (see equations (5.7.16) or (7.4.17)). In this section, we extend the AFT model, given in equation (5.7.16), to incorporate pre-treatment covariates $X_i = \mathbf{x}$. We do likewise for the censoring time component (Section 7.4.2).

7.4.1 The AFT model for failure time

Given equation (5.7.16), derived in Section 5.7, it is relatively straightforward to introduce pre-treatment covariates into the AFT model structure. The full range of possibilities for modelling covariate effects, including interactions between covariates, compliance type and non-linear covariate effects, can be considered for both the causal and selection parts of the model. The simplest model would assume no interaction between compliance type and covariates on the effect of assigned treatment, and assume that the effect of covariates on failure time, if not assigned to treatment, is the same for all compliance types. Thus the simplest model could be written as:

$$\begin{aligned} T_i(1) &= T_i(0)\exp(\zeta + \zeta^{(c)}U_i^{(c)} + \zeta^{(a)}U_i^{(a)} + \zeta_{\mathbf{x}}X_i); \\ T_i(0) &= T_{0i}\exp(\phi^{(c)}U_i^{(c)} + \phi^{(a)}U_i^{(a)} + \phi_{\mathbf{x}}X_i). \\ T_{0i} &\sim D(\theta^{(1)}, \theta^{(2)}). \end{aligned} \tag{7.4.1}$$

Let us consider another scenario. If it is necessary to explore the possibility that the effect of treatment, in particular for compliance types $U_i = u$, depends on the covariate levels $X_i = \mathbf{x}$ (e.g., the effect of treatment for compliers could depend on age, with the effectiveness of the treatment declining effects with increased age) then interaction terms need to be introduced into the casual model to relate to potential outcomes. Thus a model allowing the causal effect of treatment (i.e. assigned treatment) for particular compliance types, to vary with covariate levels can be specified as follows:

$$T_i(1) = T_i(0)\exp(\zeta + \zeta^{(c)}U_i^{(c)} + \zeta^{(a)}U_i^{(a)} + \zeta_{\mathbf{x}}X_i + \zeta_{\mathbf{x}}^{(c)}X_iU_i^{(c)} + \zeta_{\mathbf{x}}^{(a)}X_iU_i^{(a)}). \tag{7.4.2}$$

Given compliance type $U_i = u$, for $u\{c, a, n\}$, the failure time $T_i(1)$ corresponding to potential treatment level $v = 1$, can be expressed as follows:

$$T_i(1) = \begin{cases} T_i(0)\exp(\zeta + \zeta^{(c)} + (\zeta_{\mathbf{x}} + \zeta_{\mathbf{x}}^{(c)})X_i) & \text{for } U_i = c, \\ T_i(0)\exp(\zeta + \zeta^{(a)} + (\zeta_{\mathbf{x}} + \zeta_{\mathbf{x}}^{(a)})X_i) & \text{for } U_i = a, \\ T_i(0)\exp(\zeta + \zeta_{\mathbf{x}}X_i) & \text{for } U_i = n. \end{cases} \tag{7.4.3}$$

Under the assumption of *exclusion restriction*, $T_i(1) = T_i(0)$ (for $U_i = a$

or n) holds. This implies the following:

$$\begin{aligned} &\exp(\zeta + \zeta^{(a)} + (\zeta_{\mathbf{x}} + \zeta_{\mathbf{x}}^{(a)})X_i) = 1 \qquad \text{for all } X_i, \\ &\exp(\zeta + \zeta_{\mathbf{x}} X_i) = 1 \qquad\qquad\qquad\qquad \text{for all } X_i, \end{aligned} \tag{7.4.4}$$

which are equivalent to

$$\begin{aligned} \zeta + \zeta_{\mathbf{x}} X_i = 0 \qquad &\text{for all } X_i, \\ \zeta^{(a)} + \zeta_{\mathbf{x}}^{(a)} X_i = 0 \qquad &\text{for all } X_i. \end{aligned} \tag{7.4.5}$$

Equation (7.4.3) can then be simplified as follows:

$$T_i(1) = \begin{cases} T_i(0)\exp(\zeta^{(c)} + \zeta_{\mathbf{x}}^{(c)} X_i) & \text{for } U_i = c, \\ T_i(0) & \text{for } U_i = a, \\ T_i(0) & \text{for } U_i = n. \end{cases} \tag{7.4.6}$$

Equation (7.4.2) is then easily reduced to

$$T_i(1) = T_i(0)\exp(\zeta^{(c)} U_i^{(c)} + \zeta_{\mathbf{x}}^{(c)} X_i U_i^{(c)}). \tag{7.4.7}$$

Similarly, interaction terms can be introduced into the model for failure time, in the absence of treatment (i.e. if not assigned to the treatment group), if the differences between compliance types U_i are likely to vary depending on covariate level. We then have the following formulation:

$$T_i(0) = T_{0i}\exp(\phi^{(c)} U_i^{(c)} + \phi^{(a)} U_i^{(a)} + \phi_{\mathbf{x}} X_i + \phi_{\mathbf{x}}^{(c)} X_i U_i^{(c)} + \phi_{\mathbf{x}}^{(a)} X_i U_i^{(a)}), \tag{7.4.8}$$

which is equivalent to:

$$T_i(0) = \begin{cases} T_{0i}\exp(\phi^{(c)} + \phi_{\mathbf{x}} X_i + \phi_{\mathbf{x}}^{(c)} X_i) & \text{for } U_i = c, \\ T_{0i}\exp(\phi^{(a)} + \phi_{\mathbf{x}} X_i + \phi_{\mathbf{x}}^{(a)} X_i) & \text{for } U_i = a, \\ T_{0i}\exp(\phi_{\mathbf{x}} X_i) & \text{for } U_i = n. \end{cases} \tag{7.4.9}$$

Combining equations (7.4.7) and (7.4.8), we have:

$$\begin{aligned}
&T_i(1)\\
=&\left\{\begin{array}{l} T_{0i}\exp(\phi^{(c)}U_i^{(c)}+\phi^{(a)}U_i^{(a)}+\phi_{\mathbf{x}}X_i+\phi_{\mathbf{x}}^{(c)}X_iU_i^{(c)}+\phi_{\mathbf{x}}^{(a)}X_iU_i^{(a)}) \\ \exp(\zeta^{(c)}U_i^{(c)}+\zeta_{\mathbf{x}}^{(c)}X_iU_i^{(c)}) \end{array}\right\},\\
=&T_{0i}\exp((\phi^{(c)}+\zeta^{(c)})U_i^{(c)}+\phi^{(a)}U_i^{(a)}+\phi_{\mathbf{x}}X_i+(\phi_{\mathbf{x}}^{(c)}+\zeta_{\mathbf{x}}^{(c)})X_iU_i^{(c)}+\phi_{\mathbf{x}}^{(a)}X_iU_i^{(a)}).
\end{aligned} \tag{7.4.10}$$

Given compliance type $U_i = u$, (for $u \in \{c,\ a,\ n\}$), equation (7.4.10) can then be specified as the following relationships:

$$T_i(1) = \left\{\begin{array}{ll} T_{0i}\exp(\zeta^{(c)}+\phi^{(c)}+(\phi_{\mathbf{x}}+\phi_{\mathbf{x}}^{(c)})X_i) & \text{for } U_i = c, \\ T_{0i}\exp(\phi^{(a)}+(\phi_{\mathbf{x}}+\phi_{\mathbf{x}}^{(a)})X_i) & \text{for } U_i = a, \\ T_{0i}\exp(\phi_{\mathbf{x}}X_i) & \text{for } U_i = n. \end{array}\right. \tag{7.4.11}$$

Similarly, equation (7.4.8), is specified as follows:

$$T_i(0) = \left\{\begin{array}{ll} T_{0i}\exp(\phi^{(c)}+(\phi_{\mathbf{x}}+\phi_{\mathbf{x}}^{(c)})X_i) & \text{for } U_i = c, \\ T_{0i}\exp(\phi^{(a)}+(\phi_{\mathbf{x}}+\phi_{\mathbf{x}}^{(a)})X_i) & \text{for } U_i = a, \\ T_{0i}\exp(\phi_{\mathbf{x}}X_i) & \text{for } U_i = n. \end{array}\right. \tag{7.4.12}$$

Suppose that in a RCT, P covariates are involved . In this case, X_i denotes a column vector of length P,

$$X_i = (X_i^{(1)}, X_i^{(2)}, \ldots, X_i^{(P)}).^T$$

Hence its associated regression coefficients, $\phi_{\mathbf{x}}$ can be viewed as a row vector with length P, i.e:

$$\phi_{\mathbf{x}} = (\phi_{\mathbf{x}}^{(1)}, \phi_{\mathbf{x}}^{(2)}, \ldots, \phi_{\mathbf{x}}^{(P)}).$$

This implies that $\phi_{\mathbf{x}}X_i$ can be expanded as

$$\sum_{p=1}^{P}\phi_{\mathbf{x}}^{(p)}X_i^{(p)}, \tag{7.4.13}$$

if these P covariates involved in the RCT are continuous or categorical (with only two levels) variables. When the categorical covariates have more than

two levels, then more dummy variables are required[1].

Similarly, the coefficients associated with the interaction term of covariates X_i and compliance type U_i, which we denote by $\phi_{\mathbf{x}}^{(c)}$, $\phi_{\mathbf{x}}^{(a)}$, $\zeta_x^{(c)}$ and $\zeta_x^{(c)}$, can all be written as row vectors with length P. These can be expressed as:

$$\begin{aligned}
\phi_{\mathbf{x}}^{(c)} &= (\phi_{\mathbf{x}}^{(c1)}, \phi_{\mathbf{x}}^{(c2)}, \ldots, \phi_{\mathbf{x}}^{(cp)}); \\
\phi_{\mathbf{x}}^{(a)} &= (\phi_{\mathbf{x}}^{(a1)}, \phi_{\mathbf{x}}^{(a2)}, \ldots, \phi_{\mathbf{x}}^{(ap)}); \\
\zeta_{\mathbf{x}}^{(c)} &= (\zeta_{\mathbf{x}}^{(c1)}, \zeta_{\mathbf{x}}^{(c2)}, \ldots, \zeta_{\mathbf{x}}^{(cp)}).
\end{aligned} \tag{7.4.15}$$

The term $\phi_{\mathbf{x}}^{(u)} X_i$ (for $u = c$ or a) can also be written as

$$\begin{aligned}
\sum_{p=1}^{P} \phi_{\mathbf{x}}^{(cp)} X_i^{(p)} &= \phi_{\mathbf{x}}^{(c1)} X_i^{(1)} + \phi_{\mathbf{x}}^{(c2)} X_i^{(2)} + \ldots + \phi_{\mathbf{x}}^{(cP)} X_i^{(P)}; \\
\sum_{p=1}^{P} \phi_{\mathbf{x}}^{(ap)} X_i^{(p)} &= \phi_{\mathbf{x}}^{(a1)} X_i^{(1)} + \phi_{\mathbf{x}}^{(a2)} X_i^{(2)} + \ldots + \phi_{\mathbf{x}}^{(aP)} X_i^{(P)}; \\
\sum_{p=1}^{P} \zeta_{\mathbf{x}}^{(cp)} X_i^{(p)} &= \zeta_{\mathbf{x}}^{(c1)} X_i^{(1)} + \phi_{\mathbf{x}}^{(c2)} X_i^{(2)} + \ldots + \zeta_{\mathbf{x}}^{(cP)} X_i^{(P)}.
\end{aligned} \tag{7.4.16}$$

Note that the forumulae given in equations (7.4.13) and (7.4.16) are only for the cases when the covariates involved are continuous variables or categorical variables with only two levels. When categorical covariates have more than two levels, more dummy variables are required. Consequently, the number of regression coefficients will increase. For notational convenience, we continue to use $\phi_{\mathbf{x}} X_i$ to denote the regression terms in the AFT model rather than using their expansions as given in equation (7.4.13). However,

---

[1]More generally, when the *pth* pre-treatment covariate variable involved in the RCT, $X_i^{(p)}$, has more than two levels, $l_p > 2$, we can then use l_p-1 dummy variables

$$X_i^{(p1)}, X_i^{(p2)}, \ldots, X_i^{(p(1_p-1))}$$

to replace $X_i^{(p)}$. Hence the term $\phi_{\mathbf{x}}^{(p)} X_i^{(p)}$ can then be replaced by:

$$\sum_{l=1}^{l_p-1} \phi_{\mathbf{x}}^{(pl)} X_i^{(pl)} = \phi_{\mathbf{x}}^{(p1)} X_i^{(p1)} + \zeta_{\mathbf{x}}^{(p2)} X_i^{(p2)} + \ldots + \phi_{\mathbf{x}}^{(p(l_p-1))} X_i^{(p(l_p-1))}. \tag{7.4.14}$$

we need to keep in mind that $\phi_{\mathbf{x}} X_i$ can be replaced by $\sum_{p=1} \sum_{l=1}^{l_p-1} \phi_{\mathbf{x}}^{(pl)} X_i^{(pl)}$ where l_p denotes the *pth* covariate's categorical level and $l_p - 1$ is the number of dummy variables involved in the expansion for the *pth* covariate.

7.4.2 The AFT model for censoring time

Similarly, for censoring time, its AFT model extension which incorporates pretreatment covariates X_i, can be formulated as follows:

$$\begin{aligned} & C_i(1) \\ =& \left\{ \begin{array}{l} C_{0i} \exp(\varphi^{(c)} U_i^{(c)} + \varphi^{(a)} U_i^{(a)} + \varphi_{\mathbf{x}} X_i + \varphi_{\mathbf{x}}^{(c)} X_i U_i^{(c)} + \varphi_{\mathbf{x}}^{(a)} X_i U_i^{(a)}) \\ \exp(\xi^{(c)} U_i^{(c)} + \xi_{\mathbf{x}}^{(c)} X_i U_i^{(c)}); \end{array} \right\} \\ =& C_{0i} \exp((\varphi^{(c)} + \xi^{(c)}) U_i^{(c)} + \varphi^{(a)} U_i^{(a)} + \varphi_{\mathbf{x}} X_i + (\varphi_{\mathbf{x}}^{(c)} + \xi_{\mathbf{x}}^{(c)}) X_i U_i^{(c)} + \varphi_{\mathbf{x}}^{(a)} X_i U_i^{(a)}); \end{aligned} \tag{7.4.17}$$

This is because under the assumption of *exclusion restriction* (Angrist et al., 1996; Frangakis & Rubin, 1999), we have:

$$\begin{aligned} \xi + \xi_{\mathbf{x}} X_i = 0 \qquad & \text{for all } X_i, \\ \xi^{(a)} + \xi_{\mathbf{x}}^{(a)} X_i = 0 \qquad & \text{for all } X_i. \end{aligned} \tag{7.4.18}$$

(See Section 7.4.1).

According to equation (7.4.17), the AFT model for censoring time, allowing pre-treatment covariates $X_i = \mathbf{x}$, can be expressed as:

$$\begin{aligned} C_i(1) &= \left\{ \begin{array}{ll} C_{0i} \exp(\xi^{(c)} + \varphi^{(c)} + (\varphi_{\mathbf{x}} + \varphi_{\mathbf{x}}^{(c)}) X_i) & \text{for } U_i = c, \\ C_{0i} \exp(\varphi^{(a)} + (\varphi_{\mathbf{x}} + \varphi_{\mathbf{x}}^{(a)}) X_i) & \text{for } U_i = a, \\ C_{0i} \exp(\varphi_{\mathbf{x}} X_i) & \text{for } U_i = n, \end{array} \right. \\ C_i(0) &= \left\{ \begin{array}{ll} C_{0i} \exp(\varphi^{(c)} + (\varphi_{\mathbf{x}} + \varphi_{\mathbf{x}}^{(c)}) X_i) & \text{for } U_i = c, \\ C_{0i} \exp(\varphi^{(a)} + (\varphi_{\mathbf{x}} + \varphi_{\mathbf{x}}^{(a)}) X_i) & \text{for } U_i = a, \\ C_{0i} \exp(\varphi_{\mathbf{x}} X_i) & \text{for } U_i = n. \end{array} \right. \end{aligned} \tag{7.4.19}$$

7.5 Model specifications

Recall that when the failure and censoring times, corresponding to the baseline, T_{0i} and C_{0i}, are assumed to follow a Weibull or a log-normal distribution,

the failure time $T_i(1)$ and censoring time $C_i(1)$, which correspond to compliance type $U_i = u$ (for $u \in \{c, a, n\}$) and covariate values $X_i = \mathbf{x}$, also follow a Weibull or a log-normal distribution. So do $T_i(0)$ and $C_i(0)$ (Chapter 5). Their relevant parametrisation can be specified as in Section 7.5.1 and Section 7.5.2.

7.5.1 Weibull specific model parameters

When the failure time, corresponding to the baseline T_{0i}, is assumed to follow a Weibull distribution with ρ_T^0 and α_T^0, we then have the NIGN-WW(II), in which the components of $\boldsymbol{\theta}_{T,uzx}$ in equation (7.3.12) can be further partitioned into two scalars. Thus the NIGN-WW(II) model involves the following parameters:

$$\begin{aligned}
\boldsymbol{\theta}_{T,c1x} &= (\rho_{T,c1x}, \alpha_{T,c1x}),\\
\boldsymbol{\theta}_{T,c0x} &= (\rho_{T,c0x}, \alpha_{T,c0x}),\\
\boldsymbol{\theta}_{T,nx} &= (\rho_{T,nx}, \alpha_{T,nx}),\\
\boldsymbol{\theta}_{T,ax} &= (\rho_{T,ax}, \alpha_{T,ax}),
\end{aligned} \tag{7.5.1}$$

where ρ_* and α_* denote the scale and shape parameter of the Weibull distribution ($_*$ indicates the subscript of the parameters).

Given the AFT model for failure time (equation (7.4.1)) and following Theorem 1, we have the following relationships between the scale parameter ρ_* of the Weibull distribution and the regression coefficients ζ_* and ϕ_* in the AFT model ($_*$ indicates the subscript of the parameters). These relationships are specified below:

$$\begin{aligned}
\rho_{T,c1x} &= \rho_T^0 \exp(\zeta^{(c)} + \phi^{(c)} + (\phi_{\mathbf{x}} + \zeta_{\mathbf{x}}^{(c)} + \phi_{\mathbf{x}}^{(c)})X_i),\\
\rho_{T,c0x} &= \rho_T^0 \exp(\phi^{(c)} + (\phi_{\mathbf{x}} + \phi_{\mathbf{x}}^{(c)})X_i),\\
\rho_{T,nx} &= \rho_T^0 \exp(\phi_{\mathbf{x}} X_i),\\
\rho_{T,ax} &= \rho_T^0 \exp(\phi^{(a)} + (\phi_{\mathbf{x}} + \phi_{\mathbf{x}}^{(a)})X_i),
\end{aligned} \tag{7.5.2}$$

and

$$\alpha_{T,c1x} = \alpha_{T,c0x} = \alpha_{T,nx} = \alpha_{T,ax} = \alpha_T^0. \tag{7.5.3}$$

Note that equation (7.5.2) is equivalent to the following:

$$\begin{aligned}
\rho_{T,c1x} &= \rho_T^0 \exp(\zeta^{(c)} + \phi^{(c)}) \exp((\phi_{\mathbf{x}} + \zeta_{\mathbf{x}}^{(c)} + \phi_{\mathbf{x}}^{(c)})X_i), \\
\rho_{T,c0x} &= \rho_T^0 \exp(\phi^{(c)}) \exp((\phi_{\mathbf{x}} + \phi_{\mathbf{x}}^{(c)}X_i), \\
\rho_{T,nx} &= \rho_T^0 \exp(\phi_{\mathbf{x}}X_i), \\
\rho_{T,ax} &= \rho_T^0 \exp(\phi^{(a)}) \exp((\phi_{\mathbf{x}} + \phi_{\mathbf{x}}^{(a)})X_i).
\end{aligned} \tag{7.5.4}$$

According to the relationships given in equation (5.8.1), we can then re-write equation (7.5.2) as follows:

$$\begin{aligned}
\rho_{T,c1x} &= \rho_{T,c1} \exp((\phi_{\mathbf{x}} + \zeta_{\mathbf{x}}^{(c)} + \phi_{\mathbf{x}}^{(c)})X_i), \\
\rho_{T,c0x} &= \rho_{T,c0} \exp((\phi_{\mathbf{x}} + \phi_{\mathbf{x}}^{(c)})X_i), \\
\rho_{T,nx} &= \rho_{T,n} \exp(\phi_{\mathbf{x}}X_i), \\
\rho_{T,ax} &= \rho_{T,a} \exp((\phi_{\mathbf{x}} + \phi_{\mathbf{x}}^{(a)})X_i).
\end{aligned} \tag{7.5.5}$$

Similarly, the parameters associated with the censoring time distribution, $\boldsymbol{\theta}_{C,uzx} = (\boldsymbol{\theta}_{C,c1}, \boldsymbol{\theta}_{C,c0}, \boldsymbol{\theta}_{C,n}\, \boldsymbol{\theta}_{C,a})$, can also be specified as follows:

$$\begin{aligned}
\boldsymbol{\theta}_{C,c1x} &= (\rho_{C,c1x}, \alpha_{C,c1x}), \\
\boldsymbol{\theta}_{C,c0x} &= (\rho_{C,c0x}, \alpha_{C,c0x}), \\
\boldsymbol{\theta}_{C,nx} &= (\rho_{C,nx}, \alpha_{C,nx}), \\
\boldsymbol{\theta}_{C,ax} &= (\rho_{C,ax}, \alpha_{C,ax}),
\end{aligned} \tag{7.5.6}$$

where $\rho_{C,*}$ and $\alpha_{C,*}$ denote the scale and shape parameter of the Weibull distribution ($C, *$ represents the subscript of ρ and α associated with the distribution of censoring time).

Furthermore, according to Theorem 1 and equation (7.4.17), these scale parameters satisfy the following relationships:

$$\begin{aligned}
\rho_{C,c1x} &= \rho_C^0 \exp(\xi^{(c)} + \varphi^{(c)} + (\xi_{\mathbf{x}}^{(c)} + \varphi_{\mathbf{x}} + \varphi_{\mathbf{x}}^{(c)})X_i), \\
\rho_{C,c0x} &= \rho_C^0 \exp(\varphi^{(c)} + (\varphi_{\mathbf{x}}^{(c)} + \varphi_{\mathbf{x}})X_i), \\
\rho_{C,nx} &= \rho_C^0 \exp(\varphi_{\mathbf{x}}X_i); \\
\rho_{C,ax} &= \rho_C^0 \exp(\varphi^{(a)} + (\varphi_{\mathbf{x}} + \varphi_{\mathbf{x}}^{(a)})X_i),
\end{aligned} \tag{7.5.7}$$

and

$$\alpha_{C,c1x} = \alpha_{C,c0x} = \alpha_{C,nx} = \alpha_{C,ax} = \alpha_C^0. \tag{7.5.8}$$

The components of equation (7.5.7) are thus equivalent to:

$$\begin{aligned}
\rho_{C,c1x} &= \rho_C^0 \exp(\xi^{(c)} + \varphi^{(c)}) \exp((\xi_{\mathbf{x}}^{(c)} + \varphi_{\mathbf{x}} + \varphi_{\mathbf{x}}^{(c)})X_i), \\
\rho_{C,c0x} &= \rho_C^0 \exp(\varphi^{(c)}) \exp((\varphi_{\mathbf{x}}^{(c)} + \varphi_{\mathbf{x}})X_i), \\
\rho_{C,nx} &= \rho_C^0 \exp(\varphi_{\mathbf{x}} X_i), \\
\rho_{C,ax} &= \rho_C^0 \exp(\varphi^{(a)}) \exp((\varphi_{\mathbf{x}} + \varphi_{\mathbf{x}}^{(a)})X_i).
\end{aligned} \tag{7.5.9}$$

Note that ρ_C^0 and α_C^0 denote the scale and shape parameter of the Weibull distribution for censoring time C_{0i} for the baseline.

After substituting the relationships given in equation (5.8.2) into equation (7.5.9), the latter can be re-written as follows:

$$\begin{aligned}
\boldsymbol{\rho}_{C,c1x} &= \rho_{C,c1} \exp((\xi_{\mathbf{x}}^{(c)} + \varphi_{\mathbf{x}} + \varphi_{\mathbf{x}}^{(c)})X_i), \\
\boldsymbol{\rho}_{C,c0x} &= \rho_{C,c0} \exp((\varphi_{\mathbf{x}}^{(c)} + \varphi_{\mathbf{x}})X_i), \\
\boldsymbol{\rho}_{C,nx} &= \rho_{C,n} \exp(\varphi_{\mathbf{x}} X_i), \\
\boldsymbol{\rho}_{C,ax} &= \rho_{C,a} \exp((\varphi_{\mathbf{x}} + \varphi_{\mathbf{x}}^{(a)})X_i).
\end{aligned} \tag{7.5.10}$$

We have now linked the two scale parameters $\boldsymbol{\rho}_{T,uz}$, $\boldsymbol{\rho}_{C,uz}$ (for the NIGN-WW model) and $\boldsymbol{\rho}_{T,uzx}$, $\boldsymbol{\rho}_{C,uzx}$ (for the NIGN-WW(II)). In other words, equations (7.5.5), (7.5.3), (7.5.10) and (7.5.8) show that we have successfully extended the NIGN-WW model to accommodate pretreatment covariates. Table 7.1 gives the relevant parameters for the extended NIGN-WW model, namely the NIGN-WW(II) model.

Table 7.1: Non-ignorable PPO Weibull survival model (II) parameters.

	Parameters in NIGN-WW (II) model
failure time	$(\rho_{T,c1}, \rho_{T,c0}, \rho_{T,n}, \rho_{T,a}, \alpha_T)$
censoring time	$(\rho_{C,c1}, \rho_{C,c0}, \rho_{C,n}, \rho_{C,a}, \alpha_C)$
AFT model	$(\zeta_{\mathbf{x}}^{(c)}, \phi_{\mathbf{x}}, \phi_{\mathbf{x}}^{(c)}, \phi_{\mathbf{x}}^{(a)}, \xi_{\mathbf{x}}^{(c)}, \varphi_{\mathbf{x}}, \varphi_{\mathbf{x}}^{(c)}, \varphi_{\mathbf{x}}^{(a)})$
Logistic model	$(\eta_{c,110}, \eta_{c,111}, \eta_{c,000}, \eta_{c,001})$

7.5.2 Log-normal specific model parameters

Similarly, when the failure time corresponding to the baseline is assumed to follow a log-normal distribution, we then have the NIGN-LL(II) model, which involves the following parameters:

$$\begin{aligned}\boldsymbol{\theta}_{T,c1x} &= (\mu_{T,c1x}, \sigma_{T,c1x});\\ \boldsymbol{\theta}_{T,c0x} &= (\mu_{T,c0x}, \sigma_{T,c0x});\\ \boldsymbol{\theta}_{T,nx} &= (\mu_{T,nx}, \sigma_{T,nx});\\ \boldsymbol{\theta}_{T,ax} &= (\mu_{T,ax}, \sigma_{T,ax});\end{aligned} \tag{7.5.11}$$

where $\mu_{T,*}$ and $\sigma_{T,*}$ denote the mean and variance parameter of the logarithm of failure time, respectively. $(T, *$ represents the related subscript of the parameters associated with faliure time distribution).

Given equation (7.4.10) and following Theorem 1, these parameters satisfy the following equalities:

$$\begin{aligned}\mu_{T,c1x} &= \mu_T^0 + \zeta^{(c)} + \phi^{(c)} + (\phi_{\mathbf{x}} + \zeta_{\mathbf{x}}^{(c)} + \phi_{\mathbf{x}}^{(c)})X_i;\\ \mu_{T,c0x} &= \mu_T^0 + \phi^{(c)} + (\phi_{\mathbf{x}} + \phi_{\mathbf{x}}^{(c)})X_i;\\ \mu_{T,nx} &= \mu_T^0 + \phi_{\mathbf{x}}X_i;\\ \mu_{T,ax} &= \mu_T^0 + \phi^{(a)} + (\phi_{\mathbf{x}} + \phi_{\mathbf{x}}^{(a)})X_i;\end{aligned} \tag{7.5.12}$$

and

$$\sigma_{T,c1x} = \sigma_{T,c0x} = \sigma_{T,nx} = \sigma_{T,ax} = \sigma_T^0. \tag{7.5.13}$$

Given the relationship between $\mu_{T,uz}$ and ζ and ϕ in equation (5.8.5), equation (7.5.12) can then be rewritten as follows:

$$\begin{aligned}\mu_{T,c1x} &= \mu_{T,c1} + (\phi_{\mathbf{x}} + \zeta_{\mathbf{x}}^{(c)} + \phi_{\mathbf{x}}^{(c)})X_i;\\ \mu_{T,c0x} &= \mu_{T,c0} + (\phi_{\mathbf{x}} + \phi_{\mathbf{x}}^{(c)})X_i;\\ \mu_{T,nx} &= \mu_{T,n} + (\phi_{\mathbf{x}})X_i;\\ \mu_{T,ax} &= \mu_{T,a} + (\phi_{\mathbf{x}} + \phi_{\mathbf{x}}^{(a)})X_i.\end{aligned} \tag{7.5.14}$$

Similarly, for the censoring time, the following parameters can also per-

tain:

$$\begin{aligned}
\mu_{C,c1x} &= \mu_{C,c1} + (\varphi_{\mathbf{x}} + \xi_{\mathbf{x}}^{(c)} + \varphi_{\mathbf{x}}^{(c)})X_i; \\
\mu_{C,c0x} &= \mu_{C,c0} + (\varphi_{\mathbf{x}} + \varphi_{\mathbf{x}}^{(c)})X_i; \\
\mu_{C,nx} &= \mu_{C,n} + (\varphi_{\mathbf{x}})X_i; \\
\mu_{C,ax} &= \mu_{C,a} + (\varphi_{\mathbf{x}} + \varphi_{\mathbf{x}}^{(a)})X_i;
\end{aligned} \tag{7.5.15}$$

and

$$\sigma_{C,c1x} = \sigma_{C,c0x} = \sigma_{C,nx} = \sigma_{C,ax} = \sigma_C^0. \tag{7.5.16}$$

Consequently, the parameters of the log-normal PPO-survival model(s)(II) for the NIGN-LL(II) survival model, are as given in Table 7.2.

Table 7.2: Non-ignorable PPO log-normal survival model (II) parameters.

	Parameters in NIGN-LL (II) model
failure time	$(\mu_{T,c1}, \mu_{T,c0}, \mu_{T,n}, \mu_{T,a}, \sigma_T)$
censoring time	$(\mu_{C,c1}, \mu_{C,c0}, \mu_{C,n}, \mu_{C,a}, \sigma_C)$
AFT model	$(\zeta_{\mathbf{x}}^{(c)}, \phi_{\mathbf{x}}, \phi_{\mathbf{x}}^{(c)}, \phi_{\mathbf{x}}^{(a)}, \xi_{\mathbf{x}}^{(c)}, \varphi_{\mathbf{x}}, \varphi_{\mathbf{x}}^{(c)}, \varphi_{\mathbf{x}}^{(a)})$
Logistic model	$(\eta_{c,110}, \eta_{c,111}, \eta_{c,000}, \eta_{c,001})$

7.6 Computational issues

The extended model, including covariates, can again be fitted using the EM algorithm, alternating between maximising the expected log-likelihood and computing the posterior compliance type probabilities.

7.6.1 Posterior probability of compliance type U_i

Let P_{acx}^i, P_{ncx}^i, P_{nx}^i, P_{ax}^i and Q_{acx}^i, Q_{ncx}^i, Q_{nx}^i, Q_{ax}^i denote the posterior probabilities for non-censored ($R_i = 1$) and censored ($R_i = 0$) individuals, respectively, for the following cases: (1) $Z_i = 1$, $D_i(1) = 1$ and $X_i = \mathbf{x}$; (2) $Z_i = 0$, $D_i(0) = 0$ and $X_i = \mathbf{x}$; (3) $Z_i = 1$, $D_i(1) = 0$ and $X_i = \mathbf{x}$; and (4) $Z_i = 0$, $D_i(0) = 1$ and $X_i = x$.

These posterior probabilities can then be expressed as follows:

$$\begin{aligned}
P^i_{acx} &= Pr(U_i = c \,|\, (T_i(1),\, C_i(1))^{obs},\, Z_i = 1,\, D_i(1) = 1,\, X_i = x;\, \boldsymbol{\beta}^k) && \text{for } R_i = 1,\\
P^i_{ncx} &= Pr(U_i = c \,|\, (T_i(0),\, C_i(0))^{obs},\, Z_i = 0,\, D_i(0) = 0,\, X_i = x;\, \boldsymbol{\beta}^k) && \text{for } R_i = 1,\\
P^i_{nx} &= Pr(U_i = n \,|\, (T_i(1),\, C_i(1))^{obs},\, Z_i = 1,\, D_i(1) = 0,\, X_i = x;;\, \boldsymbol{\beta}^k) && \text{for } R_i = 1,\\
P^i_{ax} &= Pr(U_i = a \,|\, (T_i(0),\, C_i(0))^{obs},\, Z_i = 0,\, D_i(0) = 1,\, X_i = x;\, \boldsymbol{\beta}^k) && \text{for } R_i = 1,\\
Q^i_{acx} &= Pr(U_i = c \,|\, (T_i(1),\, C_i(1))^{obs},\, Z_i = 1,\, D_i(1) = 1,\, X_i = x;\, \boldsymbol{\beta}^k) && \text{for } R_i = 0,\\
Q^i_{ncx} &= Pr(U_i = c \,|\, (T_i(0),\, C_i(0))^{obs},\, Z_i = 0,\, D_i(0) = 0,\, X_i = x;\, \boldsymbol{\beta}^k) && \text{for } R_i = 0,\\
Q^i_{nx} &= Pr(U_i = n \,|\, (T_i(1),\, C_i(1))^{obs},\, Z_i = 1,\, D_i(1) = 0,\, X_i = x;\, \boldsymbol{\beta}^k) && \text{for } R_i = 0,\\
Q^i_{ax} &= Pr(U_i = a \,|\, (T_i(0),\, C_i(0))^{obs},\, Z_i = 0,\, D_i(0) = 1,\, X_i = x;\, \boldsymbol{\beta}^k) && \text{for } R_i = 0.
\end{aligned} \tag{7.6.1}$$

where $(T_i, C_i)^{obs}$ is as given in equation (5.5.32), which indicates the observable part of the survival outcomes $(T_i,\ C_i)$. $(T_i, C_i)^{obs}$ is equivalent to $T_i^{obs} = T_i(z)$ and $C_i^{obs} > T_i(z)$ given $Z_i = z$, when $R_i = 1$; and $(T_i, C_i)^{obs}$ is equivalent to $C_i^{obs} = C_i(z)$ and $T_i^{obs} > C_i(z)$ given $Z_i = z$, for $R_i = 0$.

Because of the assumption of *monotonicity*, which ensures that no defiers exist (Angrist et al. (1996); Frangakis & Rubin (1999) and see Section 5.2), an individual with $Z_i = 1$ and $D_i(1) = 0$, has his/her compliance type U_i fixed at that of a *never-taker*, $U_i = n$, i.e.:

$$Pr(U_i = n \mid (T_i,\ C_i)^{obs},\ Z_i = 1,\ D_i(1) = 0;\ X_i; \boldsymbol{\beta}^k) = 1. \tag{7.6.2}$$

Other individuals with $Z_i = 0$ and $D_i(0) = 1$ have compliance type U_i which must be that of an *always-taker*, $U_i = a$. We then have:

$$Pr(U_i = a \mid (T_i,\ C_i)^{obs},\ Z_i = 0,\ D_i(0) = 1;\ X_i; \boldsymbol{\beta}^k) = 1. \tag{7.6.3}$$

According to equations (7.6.2) and (7.6.3), we have the following equivalences:

$$\begin{aligned}
P^i_{nx} &= 1 \qquad \text{for } \ i \in \mathcal{C}(1, 0, 1, \mathbf{x}),\\
Q^i_{nx} &= 1 \qquad \text{for } \ i \in \mathcal{C}(1, 0, 0, \mathbf{x}),\\
P^i_{ax} &= 1 \qquad \text{for } \ i \in \mathcal{C}(0, 1, 1, \mathbf{x}),\\
Q^i_{ax} &= 1 \qquad \text{for } \ i \in \mathcal{C}(0, 1, 0, \mathbf{x}),
\end{aligned} \tag{7.6.4}$$

where $\mathcal{C}(z, d, r, \mathbf{x})$ denotes the subset of individuals, which is defined by the following: $\mathcal{C}(z, d, r, \mathbf{x}) = \{i :\ Z_i = z,\ D_i = d,\ R_i = r,\ \mathbf{x}\}$ for all z, d, $r \in \{1, 0\}$ and $\mathbf{x} \in \mathcal{X}$.

Equation (7.6.4) then provides the required posterior probabilities for the last two cases, i.e. $Z_i = 1$, $D_i(1) = 0$ and $X_i = \mathbf{x}$, and $Z_i = 0$, $D_i(0) = 1$ and $X_i = \mathbf{x}$.

In the formulation of the posterior probabilities associated with the remaining two cases: ($Z_i = 1$, $D_i(1) = 1$ and $X_i = \mathbf{x}$, and $Z_i = 0$, $D_i(0) = 0$ and $X_i = \mathbf{x}$), we need to follow Bayes' theorem, with additional pre-treatment covariates X_i included. The posterior probabilities for the non-ignorable PPO survival model can thus be expressed as follows:

$$Pr(U_i = u \mid (T_i, C_i)^{obs},\ Z_i = 1,\ D_i(1) = 1,\ X_i = \mathbf{x}; \boldsymbol{\beta}^k)$$
$$= \begin{cases} \frac{f_{ux}(1,t;\theta_{T,u1x})G_{ux}(1,t;\theta_{C,u1x})\pi_{u,11x}}{f_{cx}(1,t;\theta_{T,c1x})G_{cx}(1,t;\theta_{C,c1x})\pi_{c,11x}+f_{ax}(1,t;\theta_{T,a1x})G_{ax}(1,t;\theta_{C,a1x})\pi_{a,11x}}, & \text{if } R_i = 1,\ T_i(1) = t, \\ \frac{S_{ux}(1,t;\theta_{T,u1x})e_{ux}(1,t;\theta_{T,u1x})\pi_{u,11x}}{S_{cx}(1,t;\theta_{T,c1x})e_{c1x}(1,t;\theta_{C,c1x})\pi_{c,11x}+S_{ax}(1,t;\theta_{T,a1x})e_{ax}(1,t;\theta_{C,a1x})\pi_{a,11x}}, & \text{if } R_i = 0,\ C_i(1) = t, \end{cases} \tag{7.6.5}$$

for $u \in \{c,\ a\}$ and $\mathbf{x} \in \mathcal{X}$; and:

$$Pr(U_i = u \mid (T_i, C_i)^{obs},\ Z_i = 0,\ D_i(0) = 0;\ X_i = \mathbf{x}; \beta^k)$$
$$= \begin{cases} \frac{f_{ux}(0,t;\theta_{T,u0x})G_{ux}(1,t;\theta_{C,u0x})\pi_{u,00x}}{f_{cx}(0,t;\theta_{T,c0x})G_{cx}(0,t;\theta_{C,c0x})\pi_{c,00x}+f_{nx}(0,t;\theta_{T,n0x})G_{nx}(0,t;\theta_{C,n0x})\pi_{n,00x}} & \text{if } R_i = 1,\ T_i(0) = t, \\ \frac{S_{ux}(0,t;\theta_{T,u0x})e_{ux}(0,t;\theta_{C,u0x})\pi_{u,00x}}{S_{cx}(0,t;\theta_{T,c0x})e_{cx}(0,t;\theta_{C,c0x})\pi_{c,00x}+S_{nx}(0,t;\theta_{T,n0x})e_{nx}(0,t;\theta_{C,n0x})\pi_{n,00x}} & \text{if } R_i = 0,\ C_i(0) = t, \end{cases} \tag{7.6.6}$$

where $u \in \{c,\ n\}$ and $\mathbf{x} \in \mathcal{X}$.

Therefore, P^i_{acx}, P^i_{ncx}, Q^i_{acx}, Q^i_{ncx} can then be formulated as in equation (7.6.7) given overleaf.

$$
\begin{aligned}
P^i_{acx} &= \frac{f_{cx}(1,t;\theta_{T,c1x})G_{cx}(1,t;\theta_{C,c1x})\pi_{c,11x}}{f_{cx}(1,t;\theta_{T,c1x})G_{cx}(1,t;\theta_{C,c1x})\pi_{c,11x} + f_{ax}(1,t;\theta_{T,a1})G_{ax}(1,t;\theta_{C,a1x})\pi_{a,11x}} \quad \text{for } i \in \mathcal{C}(1,1,1,\mathbf{x}),\ T_i(1) = t, \\
P^i_{ncx} &= \frac{f_{cx}(0,t;\theta_{T,c0x})G_{cx}(1,t;\theta_{C,c0x})\pi_{c,00x}}{f_{cx}(0,t;\theta_{T,c0x})G_{cx}(0,t;\theta_{C,c0x})\pi_{c,00x} + f_{n0x}(0,t;\theta_{T,n0x})G_{nx}(0,t;\theta_{C,n0x})\pi_{n,00x}} \quad \text{for } \ i \in \mathcal{C}(0,0,1,\mathbf{x}),\ T_i(0) = t, \\
Q^i_{acx} &= \frac{S_{cx}(1,t;\theta_{T,c1x})e_{cx}(1,t;\theta_{C,c1x})\pi_{c,11x}}{S_{cx}(1,t;\theta_{T,c1x})e_{cx}(1,t;\theta_{C,c1x})\pi_{c,11} + S_a(1,t;\theta_{T,a1x})e_{ax}(1,t;\theta_{C,a1x})\pi_{a,11x}} \quad \text{for } \ i \in \mathcal{C}(1,1,0,\mathbf{x}),\ C_i(1) = t, \\
Q^i_{ncx} &= \frac{S_{cx}(0,t;\theta_{T,c0x})e_{cx}(0,t;\theta_{C,c0x})\pi_{c,00x}}{S_{cx}(0,t;\theta_{T,c0x})e_{cx}(0,t;\theta_{C,c0x})\pi_{c,00x} + S_{nx}(0,t;\theta_{T,n0x})e_{nx}(0,t;\theta_{C,n0x})\pi_{n,00x}} \quad \text{for } \ i \in \mathcal{C}(0,0,0,\mathbf{x}),\ C_i(0) = t.
\end{aligned}
\tag{7.6.7}
$$

As in Chapter 5 (Section 5.9.2), we use $W_i(u, \boldsymbol{\beta}^k)$ to denote the posterior probability of compliance type $Pr(U_i = u|(T_i, C_i)^{obs}, Z_i = z, D_i(z) = d, X_i = \mathbf{x}; \boldsymbol{\beta}^k)$ given the observed data and parameter values for the extended non-ignorable PPO-survival model. That is:

$$W_i(u, \boldsymbol{\beta}^k) = Pr(U_i = u|(T_i, C_i)^{obs},\ Z_i = z,\ D_i(z) = d,\ X_i = \mathbf{x};\ \boldsymbol{\beta}^k), \quad (7.6.8)$$

The posterior probability $W_i(u, \boldsymbol{\beta}^k)$ of the *ith* individual can then be written as:

$$W_i(u, \boldsymbol{\beta}^k) = \begin{cases} (P^i_{acx})^{I\{u=c\}}(1 - P^i_{acx})^{I\{u=a\}} & \text{for } i \in \mathcal{C}(1, 1, 1, \mathbf{x}), \\ (P^i_{ncx})^{I\{u=c\}}(1 - P^i_{ncx})^{I\{u=n\}} & \text{for } i \in \mathcal{C}(0, 0, 1, \mathbf{x}), \\ (Q^i_{acx})^{I\{u=c\}}(1 - Q^i_{acx})^{I\{u=a\}} & \text{for } i \in \mathcal{C}(1, 1, 0, \mathbf{x}), \\ (Q^i_{ncx})^{I\{u=c\}}(1 - Q^i_{ncx})^{I\{u=n\}} & \text{for } i \in \mathcal{C}(0, 0, 0, \mathbf{x}), \\ P^i_{nx} = 1 & \text{for } i \in \mathcal{C}(1, 0, 1, \mathbf{x}), \\ Q^i_{nx} = 1 & \text{for } i \in \mathcal{C}(1, 0, 0, \mathbf{x}), \\ P^i_{ax} = 1 & \text{for } i \in \mathcal{C}(0, 1, 1, \mathbf{x}), \\ Q^i_{ax} = 1 & \text{for } i \in \mathcal{C}(0, 1, 0, \mathbf{x}), \end{cases} \quad (7.6.9)$$

where $I\{\}$ is an indicator function, and P^i_{acx}, P^i_{ncx}, Q^i_{acx}, Q^i_{ncx} denote the posterior probabilities for the cases, given by equation (7.6.7) for the non-ignorable PPO-survival model, and given by equation (7.7.4) for the ignorable PPO-survival model.

7.6.2 Structures of the expectation function $Q(\boldsymbol{\beta}, \boldsymbol{\beta}^k)$

Because covariates are included in the PPO-survival model(s) (II), the parameters, π_{c11x} and π_{c00x}, associated with the conditional probability of compliance given $Z_i = z$, $D_i(z) = d$ and $X_i = \mathbf{x}$ cannot be estimated directly by following the earlier formulations (equation (5.5.27) in Chapter 5). Equation (7.3.4) has thus linked the parameters π_{c11x} and π_{c00x} with the pre-treatment covariates $X_i = \mathbf{x}$. Unlike the scenario of the PPO-survival model(s)(I), π_{c11x} and π_{c00x} need to be involved in the likelihood function. Therefore, in the PPO-survival model(s) (II), the expectation function, is still denoted by

$Q(\boldsymbol{\beta}, \boldsymbol{\beta}^k)$. However, we need to keep in mind that the model paramter β here involves more conponents than it had in the PPO-survival model(s)(I).

Table 7.1 (on page 200) and Table 7.2 (on page 202) give all parameters of $\boldsymbol{\beta}$ for the NIGN-WW and the NIGN-LL model, repectively. Note that the parameters associated with the distributions (without pretreatment covariates) of failure time ($\theta_{T,uz}$), and of the censoring time ($\theta_{C,uz}$) (for $u \in \{c,\, a,\, n\}$ and $z = 1$ or 0) are included in, $\boldsymbol{\beta}$, the parameter vector of the non-ignorable PPO-survival model. Recall that $\theta_{T,uz}$ and $\theta_{C,uz}$ (for $u \in \{c,\, a,\, n\}$ and $z = 1$ or 0) are determined by the assumed distributions (Section 5.7). Details are given by equations (5.8.1) and (5.8.2) for the Weibull model, and by equations (5.8.5) and (5.8.6) for the log-normal model.

The expectation function $Q(\boldsymbol{\beta}, \boldsymbol{\beta}^k)$ can thus be expressed as follows:

$$
\begin{aligned}
&Q^i(\boldsymbol{\beta},\boldsymbol{\beta}^k) = \\
&\sum_{u\in\{c,n,a\}} \left\{ \begin{array}{l} Pr(U_i = u|(T_i,\, C_i)^{obs},\, Z_i = z,\, D_i(z),\, X_i; \boldsymbol{\beta}^k) \\ \log\{L^E(\boldsymbol{\beta}; u)\} \end{array} \right\},
\end{aligned} \tag{7.6.10}
$$

where $L^E(\boldsymbol{\beta}; u)$ is specified by equation (7.3.10) for the non-ignorable PPO-survival model.

Using the notation in equation (7.6.9), equation (7.6.10) can be further simplified as follows:

$$
\begin{aligned}
&Q^i(\boldsymbol{\beta},\boldsymbol{\beta}^k) \\
&\quad = \sum_{u\in\{c,n,a\}} W_i(u, \boldsymbol{\beta}^k) \log\{L_i^E(\boldsymbol{\beta}; u)\}.
\end{aligned} \tag{7.6.11}
$$

Substituting equation (7.3.10) into equation (7.6.11), $Q^i(\boldsymbol{\beta}, \boldsymbol{\beta}^k)$, the expectation function for the non-ignorable PPO-survival model (for the *ith* individual), then has the following form:

$$
\begin{aligned}
&Q^i(\boldsymbol{\beta},\boldsymbol{\beta}^k) \\
&\quad = \sum_{u\in\{c,n,a\}} W(u, \boldsymbol{\beta}^k) \log \left\{ \begin{array}{l} [f^i_{ux}(z,t;\theta_{T,uzx}) G^i_{ux}(z,t;\theta_{C,uzx})]^{R_i} \\ [S^i_{ux}(z,t;\theta_{T,uzx}) e^i_{ux} z, t; \theta_{C,uzx})]^{1-R_i} \end{array} \right\} \pi_{u,zdx}.
\end{aligned} \tag{7.6.12}
$$

Recall that $\pi_{u,zdx}$ is given in equation (7.3.4) and denotes the conditional probability of compliance type, $U_i = u$, given $Z_i = z$, $D_i = d$, and $X_i = \mathbf{x}$.

The structure of the expectation function, the so-called Q-function $Q^i(\boldsymbol{\beta}, \boldsymbol{\beta}^k)$, is maximised for some initial[2] values for all relevant parameters $\boldsymbol{\beta}^0$. We then replace β^0 in $Q^i(\boldsymbol{\beta}, \boldsymbol{\beta}^0)$ with the new updated estimate $\boldsymbol{\beta}^1$, leading to $Q^i(\boldsymbol{\beta}, \boldsymbol{\beta}^1)$. This process is repeated untill convergence.

When the parameters included in the relevant model have been estimated, the survival function corresponding to the subgroups of interest can be calculated from the assumed distributional form (such as Weibull), and ACE(t) and CACE(t) can then be obtained directly (Section 3.3.2). Methods for calculating the standard error (addressed in Section 5.9.4) are still applicable for our PPO-survival model(s) (II). We omit details here.

7.7 Ignorable censoring mechanism PPO-survival model (II)

7.7.1 The likelihood function

When the censoring mechanism is ignorable, the likelihood function for the *ith* individual, given in equation (7.3.10), can then be simplified as follows:

$$\begin{aligned} &L^E{}_{i\in\{i:Z_i=z,D_i(z)=d\}}(\boldsymbol{\beta}; u) \\ &= \left\{ \begin{array}{l} [f^i_{ux}(z,t;\theta_{T,uzx}))]^{R_i} \\ [S^i_{ux}(z,t;\theta_{T,uzx}))]^{1-R_i}\ \pi_{u,zdx}. \end{array} \right\}. \end{aligned} \tag{7.7.1}$$

Equation (7.7.1) thus provides a tractable likelihood function for complete data for the ignorable PPO-survival model (II) where

$$\boldsymbol{\beta} = (\boldsymbol{\theta}_{T,uzx}, \zeta_{\mathbf{x}^{(v)}}, \phi_{\mathbf{x}^{(v)}}, \xi_{\mathbf{x}^{(v)}}, \varphi_{\mathbf{x}^{(v)}}, \boldsymbol{\eta}).$$

7.7.2 Model specification

Similar to the non-ignorable PPO-survival model (I), we take the ignorable PPO-survival model (II) to be the IGN-W(II) model and the IGN-L(II)

[2]The initial value of the missing U_i, can be generated from a Bernoulli distribution with a given value of $\hat{\pi}_c$.

model, given by a Weibull or a log-normal distribution, respectively. Table 7.3 gives all relevant components for β, for both models.

Table 7.3: Specific ignorable PPO-survival model (II) parameters.

Model	Parameters for failure time	Parameters for covariates
IGN-W(II)	$(\rho_{T,c1}, \rho_{T,c0}, \rho_{T,n}, \rho_{T,a}, \alpha)$	$(\zeta_{\mathbf{x}}, \zeta_{\mathbf{x}}^{(c)}, \phi_{\mathbf{x}}, \phi_{\mathbf{x}}^{(c)}, \phi_{\mathbf{x}}^{(a)}, \xi_{\mathbf{x}}, \xi_{\mathbf{x}}^{(c)}, \varphi_{\mathbf{x}}, \varphi_{\mathbf{x}}^{(c)}, \varphi_{\mathbf{x}}^{(a)}, \eta_{110}, \eta_{111}, \eta_{000}, \eta_{001})$
IGN-L(II)	$(\mu_{T,c1}, \mu_{T,c0}, \mu_{T,n}, \mu_{T,a}, \sigma)$	$(\zeta_{\mathbf{x}}, \zeta_{\mathbf{x}}^{(c)}, \phi_{\mathbf{x}}, \phi_{\mathbf{x}}^{(c)}, \phi_{\mathbf{x}}^{(a)}, \xi_{\mathbf{x}}, \xi_{\mathbf{x}}^{(c)}, \varphi_{\mathbf{x}}, \varphi_{\mathbf{x}}^{(c)}, \varphi_{\mathbf{x}}^{(a)}, \eta_{110}, \eta_{111}, \eta_{000}, \eta_{001})$

7.7.3 Posterior probability of compliance type

Because no censoring time distribution is included in the likelihood function of the ignorable PPO survial model, the posterior distribution of compliance type can be formulated from equations (7.6.5) and (7.6.6). We then have

$$Pr(U_i = u|(T_i,\, C_i)^{obs},\; Z_i = 1,\; D_i = 1,\; X_i = \mathbf{x}; \boldsymbol{\beta}^k) = \begin{cases} \frac{f_{ux}(1,t;\theta_{T,u1x})\pi_{u,11x}}{f_{cx}(1,t;\theta_{T,c1x})\pi_{c,11x}+f_{ax}(1,t;\theta_{T,a1x})\pi_{a,11x}} & \text{if } R_i = 1,\, T_i(1) = t, \\ \frac{S_{ux}(1,t;\theta_{T,u1x})\pi_{u,11x}}{S_{cx}(1,t;\theta_{T,c1x})\pi_{c,11x}+S_{ax}(1,t;\theta_{T,a1x})\pi_{a,11x}} & \text{if } R_i = 0,\, C_i(1) = t, \end{cases} \tag{7.7.2}$$

for $u \in \{c,\, a\}$ and $\mathbf{x} \in \mathcal{X}$, and

$$Pr(U_i = u|(T_i,\, C_i)^{obs},\; Z_i = 0,\; D_i = 0;\; X_i = \mathbf{x}; \boldsymbol{\beta}^k) = \begin{cases} \frac{f_{ux}(0,t;\theta_{T,u0x})\pi_{u,00x}}{f_{cx}(0,t;\theta_{T,c0x})\pi_{c,00x}+f_{nx}(0,t;\theta_{T,n0x})\pi_{n,00x}} & \text{if } R_i = 1,\, T_i(0) = t, \\ \frac{S_{ux}(0,t;\theta_{T,u0x})\pi_{u,00x}}{S_{cx}(0,t;\theta_{T,c0x})\pi_{c,00x}+S_{nx}(0,t;\theta_{T,n0x})\pi_{n,00x}} & \text{if } R_i = 0,\, C_i(0) = t, \end{cases} \tag{7.7.3}$$

where for $u \in \{c,\, n\}$ and $\mathbf{x} \in \mathcal{X}$.

Therefore, for the ignorable PPO-survival model, the posterior probabilities P_{acx}^i, P_{ncx}^i, Q_{acx}^i, Q_{ncx}^i can then be written as follows:

$$
\begin{aligned}
P^i_{acx} &= \frac{f_{cx}(1,t;\theta_{T,c1x})\pi_{c,11x}}{f_{cx}(1,t;\theta_{T,c1x})\pi_{c,11x}+f_{ax}(1,t;\theta_{T,a1})\pi_{a,11x}} && \text{for } i \in \mathcal{C}(1,1,1,\mathbf{x}), T_i(1) = t,\\
P^i_{ncx} &= \frac{f_{cx}(0,t;\theta_{T,c0x})\pi_{c,00x}}{f_{cx}(0,t;\theta_{T,c0x})\pi_{c,00x}+f_{nx}(0,t;\theta_{T,n0x})\pi_{n,00x}} && \text{for } i \in \mathcal{C}(0,0,1,\mathbf{x}), T_i(0) = t,\\
Q^i_{acx} &= \frac{S_{cx}(1,t;\theta_{T,c1x})\pi_{c,11x}}{S_{cx}(1,t;\theta_{T,c1x})\pi_{c,11}+S_{ax}(1,t;\theta_{T,a1x})\pi_{a,11x}} && \text{for } i \in \mathcal{C}(1,1,0,\mathbf{x}), C_i(1) = t,\\
Q^i_{ncx} &= \frac{S_{cx}(0,t;\theta_{T,c0x})\pi_{c,00x}}{S_{cx}(0,t;\theta_{T,c0x})\pi_{c,00x}+S_{nx}(0,t;\theta_{T,n0x})\pi_{n,00x}} && \text{for } i \in \mathcal{C}(0,0,0,\mathbf{x}), C_i(0) = t.
\end{aligned}
\tag{7.7.4}
$$

7.7.4 Expectation function $Q(\boldsymbol{\beta}, \boldsymbol{\beta}^k)$

By following equation (7.7.1), the expectation function, th so-called Q-function $Q^i(\boldsymbol{\beta}, \boldsymbol{\beta}^k)$, for the ignorable PPO-survival model has the following form:

$$
\begin{aligned}
&Q^i(\boldsymbol{\beta},\boldsymbol{\beta}^k)\\
&\quad = Pr(U_i = u|(T_i,\ C_i)^{obs},\ Z_i = z,\ D_i(z),\ X_i = \mathbf{x};\boldsymbol{\beta}^k)\log\left\{L^E_i(\boldsymbol{\beta};u|\boldsymbol{\beta}^k)\right\}\\
&\quad = W(u,\boldsymbol{\beta}^k)\log\left\{f^i_{ux}(z,t;\theta_{T,uzx})^{R_i}S^i_{ux}(z,t;\theta_{T,uzx})^{1-R_i}\pi_{u,zdx}\right\},
\end{aligned}
\tag{7.7.5}
$$

where $\boldsymbol{\beta} = (\boldsymbol{\theta}_{T,uzx}, \boldsymbol{\pi}_{u,zdx})$ is given in Table 7.3.

7.7.5 Computation

Similar to the derivation of the PPO-survival model (I), the ignorable PPO-survival model (II) can be implemented directly by following the methods given in Section 7.6.

7.8 Illustration of the implementation an EM algorithm for the non-ignorable PPO-survival model (II)

The parameter vector in the non-ignorable PPO-survival model (II) has more components than that one in the non-ignorable PPO-survival model (I) for the inclusion of pre-treatment covariates. For instance, in the NIGN-WW (I), only the parameter $\boldsymbol{\theta}$, a vector with 8 components in the HIP trial, need to be estimated via the EM algorithm; whereas in the NIGN-WW (II), the parameter involved into the EM alogrithm is $\boldsymbol{\beta}$ rather than $\boldsymbol{\theta}$, and it has 16 components in the HIP trial. This is because in this case, one single covariate, the age at entry is in the model, hence the regression coefficients for

both the AFT model (equations (7.4.10) and (7.4.17)) and the logistic model (equations (7.3.3) or (7.3.4)) must be included in the model. Consequently, the process of implementing the new model developed in this chapter is much more complicated than the PPO-survival model (I) derived in Chapter 5. In this section, we demonstrate how the EM algorithm is applied to the NIGN-WW (II) via the HIP_7 data.

7.8.1 Model modification of $Q(\boldsymbol{\beta}, \boldsymbol{\beta}^k)$

Recall that in the HIP data, only two compliance types and three subgroups are involved because of the assumption that there are no *always-takers* (no access to screening in the control group). So we can use $U_i = 1$ to indicate *complier*, and $U_i = 0$ to indicate *never-taker*. Equations (7.5.5) and (7.5.10), which give the relationships between the scale parameter and regression coefficients in the AFT and the logistic model for failure time and censoring time, can then be simplified as follows:

$$\begin{aligned} \rho_{T,c1x} &= \rho_{T,c1} \exp((\phi_{\mathbf{x}} + \zeta_{\mathbf{x}}^{(c)} + \phi_{\mathbf{x}}^{(c)})X_i), \\ \rho_{T,c0x} &= \rho_{T,c0} \exp((\phi_{\mathbf{x}} + \phi_{\mathbf{x}}^{(c)})X_i), \\ \rho_{T,nx} &= \rho_{T,n} \exp(\phi_{\mathbf{x}} X_i), \end{aligned} \tag{7.8.1}$$

and

$$\begin{aligned} \rho_{C,c1x} &= \rho_{C,c1} \exp((\xi_{\mathbf{x}}^{(c)} + \varphi_{\mathbf{x}} + \varphi_{\mathbf{x}}^{(c)})X_i), \\ \rho_{C,c0x} &= \rho_{C,c0} \exp((\varphi_{\mathbf{x}}^{(c)} + \varphi_{\mathbf{x}})X_i), \\ \rho_{C,nx} &= \rho_{C,n} \exp(\varphi_{\mathbf{x}} X_i), \end{aligned} \tag{7.8.2}$$

respectively.

For the same reason, the relationships between shape parameters associated with each subgroup are also simplified as follows:

$$\alpha_{T,c1x} = \alpha_{T,c0x} = \alpha_{T,nx} = \alpha_T^0. \tag{7.8.3}$$

$$\alpha_{C,c1x} = \alpha_{C,c0x} = \alpha_{C,nx} = \alpha_C^0. \tag{7.8.4}$$

Therefore, Table 7.1 given in Section 7.5.1, which lists the model parameters in the NIGN-WW (II), is simplified in the analysis of the HIP_7 data.

It is given in Table 7.4, where the parameters associated with Subgroup 4, individuals who are *always-takers*, are excluded, as well as ($\eta_{c,110}$, $\eta_{c,111}$), the regression coefficients in the logistic model for $\pi_{c,11x}$. This is because when no *always-takers* exist in the HIP trial, $\pi_{c,11x}$ has a constant value 1 for all given $\mathbf{x}$.

Table 7.4: Non-ignorable PPO Weibull survival model (II) parameters for the HIP_7 data.

	Parameters in NIGN-WW (II) model
failure time	$(\rho_{T,c1}, \rho_{T,c0}, \rho_{T,n}, \alpha_T)$
censoring time	$(\rho_{C,c1}, \rho_{C,c0}, \rho_{C,n}, \alpha_C)$
AFT model	$(\zeta_{\mathbf{x}}^{(c)}, \phi_{\mathbf{x}}, \phi_{\mathbf{x}}^{(c)}, \xi_{\mathbf{x}}^{(c)}, \varphi_{\mathbf{x}}, \varphi_{\mathbf{x}}^{(c)})$
Logistic model	$(\eta_{c,000}, \eta_{c,001})$

Based on the formulation of the expectation function $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^k)$ associated with the NIGN-WW (I), given in equation (5.9.12), the expectation function associated with the NIGN-WW (II), denoted by $Q(\boldsymbol{\beta}, \boldsymbol{\beta}^k)$ and given in equation (7.6.11), can be modified by adding in equations (7.8.1), (7.8.2), (7.8.3) and (7.8.4). Note that the first formula in equations (7.8.1) and (7.8.2) include two regression coefficients which both correspond to the interactive term of compliance type and age covariates $X \times U$ in the AFT model. Recall that for compliers, we have $U_i = 1$, which is why compliance type U did not appear in these related regression AFT models.

7.8.2 Preparation of computation

Normalisation of covariate

To avoid extreme values of the scale parameter, the covariate X is normalised by the given formula:

$$X = (X - mean(X))/std(X) \tag{7.8.5}$$

where $mean()$ and $std()$ denote the functions of average and standard deviation in MATLAB.

Imputation of missing compliance type for initial values

The compliance type covariate can be treated as a variable with missing data since those individuals who are assigned to the control group, have unobservable compliance types. In order to initiate the EM algorithm, we impute compliance types by generating them, from a Bernoulli distribution given parameter $\hat{\pi}_c$. This is because individuals in the HIP data have only two types of compliance behaviours: *compliers* and *never-takers*. The value of $\hat{\pi}_c$ is given as 0.6698, which was obtained in Chapter 6. Note that these generated values of compliance type are only used in calculating the initial parameter values.

Initial values

The NIGM-WW(II) for the HIP_7 data involves 16 parameters in total (Table 7.4). Eight of them are scale and shape parameters corresponding to the PPO-survival Weibull model (I); these eight parameters initial values are given directly as those obtained via the NIGN-WW(I) given in (Table C-1 in Appendix C). The remainding parameters are the regression coefficients that are associated with the AFT model and logistic model.

For the AFT model, six regression coefficients are involved in the analysis of the HIP data, denoted by ζ_c, ϕ_c, ϕ_x (for failure time) and ξ_c, φ_c, φ_x (for censoring time). They correspond to the interactive term $X \times U$ with $U = 1$, and term X (age at the year of enrollment in the HIP data), also used in equations (7.8.1) and (7.8.2) for calculating the scale parameter in the Weibull distribution of failure time and censoring time for each subgroup.

We use the **glmfit** command in MATLAB to estimate the regression coefficients in the AFT model, which are only used as the initial values in our extended non-ignorable PPO-Survival model; where the required distribution of input data and link function were given as *normal* and *log*. The related outputs for the HIP_7 data in the AFT model are given in Table 7.5.

Note that the intercept in the AFT model is not included in either equation (7.8.1) or (7.8.2), so the intercept in Table 7.5 needs to be excluded from intitial values. In addition, the coefficient for the interactive term $X \times U$ is

Table 7.5: Regression coefficients in the AFT model for the HIP_7 data.

Model	AFT	Data	HIP_7
Coefficient	Intercept	X	$X \times U^a$
failure time	1.5864	-0.0187	0.0315
censoring time	1.8804	-0.0070	-0.0029

[a]U value is given by imputation in Bernoulli distribution

related to two parameters ζ_c, ϕ_c (in the AFT model for failure time) and ξ_c, φ_c (in the AFT censoring time model). Hence we give half of the estimated coefficent value of the interactive term $X \times U$ as the as the initial value of relevent parameter. Note that according to the AFT model given on p.46 of Klein (1997), equations (7.8.1) and (7.8.2) should take the negative values of parameters obtained by regression formula. Therefore, the initial values for the parameters associated with the AFT model are given in Table 7.6.

Table 7.6: Initial values for the parameters corresponding to the regression coefficients in the AFT model for the HIP_7 data.

Model	AFT			Data	$HIP_7{}^a$	
	Failure time			Censoring time		
parameter	ζ_c	ϕ_c	ϕ_x	ξ_c	φ_c	φ_x
initial value	-0.01575	-0.01575	0.0187	0.00145	0.00145	0.007

[a]U value is given by imputation in Bernoulli distribution

For the logistic model, we also use the **glmfit** command in MATLAB to find the initial values of the logistic regression coefficients by setting the distribution to binomial and using the imputed compliance types as the outcome variables. Table 7.7 lists the regression coefficients in the logistic model. By now, the initial values for analysing the HIP_7 data with the NIGN-WW (II) model are all obtained, they are given in Table 7.8.

Table 7.7: Regression coefficients in the logistic model ($\pi_{c,00x}$) for the HIP_7 data.

Model	logistic	Data: HIP_7[a]
coefficient	η_{000}	η_{001}
model fitted	0.6944	0.0020

[a]U value is given by imputation in Bernoulli distribution

Table 7.8: Initial values of model parameter β^0 in the NIGN-WW(II) model for the HIP_7 data.

Weibull model	failure time	$\rho_{T,c1}$	$\rho_{T,c0}$	$\rho_{T,n}$	α_T^0
		169.49	108.7	144.93	1.9362
	censoring time	$\rho_{C,c1}$	$\rho_{C,c0}$	$\rho_{C,n}$	α_C^0
		6.9541	6.9156	6.8871	10.378
AFT model	failure time	$\zeta_{\mathbf{x}}^{(c)}$	$\phi_{\mathbf{x}}^{(c)}$	$\phi_{\mathbf{x}}$	
		-0.01575	-0.01575	0.0187	
	censoring time	$\xi_{\mathbf{x}}^{(c)}$	$\varphi_{\mathbf{x}}^{(c)}$	$\varphi_{\mathbf{x}}$	
		0.00145	0.00145	0.007	
Logistic model		$\eta_{c,000}$	$\eta_{c,001}$		
		0.6944	0.002		

7.8.3 Implementation of EM alogrithm

The first maximisation

As in the NIGM-WW(I), the **fminsearch** command in MATLAB is used to minimise the negative Q-function $-Q(\boldsymbol{\beta}, \boldsymbol{\beta}^k)$, which is equivalent to maximising the Q-function $Q(\boldsymbol{\beta}, \boldsymbol{\beta}^k)$.

We first obtain the initial values of model parameter $\boldsymbol{\beta}$ as shown in Table 7.8(Section 7.8.2):

$$\boldsymbol{\beta}^0 = [169.49,\ 108.7,\ 144.93,\ 1.9362,\ 6.9541,\ 6.9156,\ 6.8871,\ 10.378,$$
$$-0.01575,\ -0.01575,\ 0.0187,\ 0.00145,\ 0.00145,\ 0.007,\ 0.6944,\ 0.002].$$

Then the estimated values of model parameter $\boldsymbol{\beta}^1$ are obtained by running

the NIGM-WW(II) for the HIP_7 data, where the value of the posterior probabilities of compliance type are given based on the value of $\boldsymbol{\beta}^0$.

$$\boldsymbol{\beta}^1 = [86.7594,\ 59.1747,\ 81.5242,\ 2.3221,\ 6.9525,\ 6.9614,\ 6.8703,\ 10.3718,$$
$$-0.0279,\ -0.0213,\ 0.0033,\ 0.0018,\ 0.0017,\ 0.0083,\ 0.4053,\ 0.0030];$$

Note that for the purposes of illustrating the computations, we reduced the computing time by setting the termination criteria in the maximisation both for $\boldsymbol{\beta}$ and $Q(\boldsymbol{\beta}, \boldsymbol{\beta}^k)$ at 0.1 level. According to the output messages of running **fminsearch** in MATLAB, it has taken 150 iterations to reach the maximazation of $Q(\boldsymbol{\beta}, \boldsymbol{\beta}^k)$ in this first M-step.

The iterations in the EM alogrithm

Using $\boldsymbol{\beta}^1$ as a new initial value for the model parameter in the NIGN-WW (II) corresponding to the HIP_7 data and after running the NIGN-WW (II), the estimated values of model parameter $\boldsymbol{\beta}^2$ are obtained, where the value of the posterior probabilities of compliance type are given based on the value of $\boldsymbol{\beta}^1$. After k iterations, the initial parameter is $\boldsymbol{\beta}^k$, and the new estimated model parameter is $\boldsymbol{\beta}^{k+1}$, which maximises the expectation function $Q(\boldsymbol{\beta}, \boldsymbol{\beta}^k)$. These iterations are continued until convergence is achieved.

Table 7.9 shows the following six iterations (for $k = 2$ to 7) in the EM alogrithm for the HIP_7 data, where β^1 is applied as a new inital value. From the table, we can read that $\Delta = 0.0243$ when $k = 7$. Clearly, the iteration would need to be continued in order to meet more stringent convergence criteria. However, we stop the computation here as it is enough to demonstrate how the EM algorithm is applied in the NIGN-WW (II) model.

Table 7.9: Iterations of the EM algorithm using the NIGN-WW (II) model for the HIP_7 data.

Iter k[a]	$\theta^{(1)}_{T,c1}$	$\theta^{(1)}_{T,c0}$	$\theta^{(1)}_{T,n0}$	α	$\theta^{(1)}_{C,c1}$	$\theta^{(1)}_{C,c0}$	$\theta^{(1)}_{C,n0}$	α_c	ζ_c	ϕ_c	ϕ_x	ξ_c	φ_c	φ_x	η_{000}	η_{001}	Mflag [b]	Δ^c
2	85.77	67.10	72.65	2.33	6.95	6.92	6.89	10.37	−0.0171	−0.0272	0.0132	0.0015	0.0012	0.0083	0.6041	0.0029	1	83.7250
3	85.77	65.67	73.83	2.33	6.95	6.92	6.89	10.37	−0.0173	−0.0275	0.0132	0.0016	0.0012	0.0083	0.6068	0.0029	1	1.4316
4	85.75	65.65	73.86	2.33	6.95	6.92	6.89	10.37	−0.0174	−0.0276	0.0133	0.0016	0.0012	0.0083	0.6101	0.0029	1	0.0316
5	85.77	65.66	73.89	2.33	6.95	6.92	6.89	10.37	−0.0175	−0.0278	0.0134	0.0016	0.0012	0.0084	0.6135	0.0029	1	0.0227
6	85.79	65.65	73.91	2.33	6.95	6.92	6.89	10.37	−0.0176	−0.0279	0.0135	0.0016	0.0012	0.0084	0.6169	0.0029	1	0.0245
7	85.81	65.65	73.94	2.33	6.95	6.92	6.89	10.37	−0.0177	−0.0281	0.0135	0.0016	0.0012	0.0085	0.6203	0.0029	1	0.0243

[a] β^1 is not included here, as it is given as a new initial value in the following iterations.

[b] $Mflag = 1$ indicates the maximum of $Q(\theta, \theta^{(k)})$ is reached in the k^{th} iteration.

[c] Δ represents the maximum component of $|\theta^{(k+1)} - \theta^{(k)}|$, the absolute vector of difference.

7.8.4 Conclusion

The computing process in the NIGN-WW(II) model is much more complicated than the NIGN-WW(I) as the dimension of the model parameter vector increases considerably when covariates are included in the model. As mentioned before in the HIP_7 data, only one pre-treatment covariate (the age at the year of enrollment) was added in, the dimension of model parameter in the NIGN-WW(II) is doubled as the one in the NIGN-WW(I),(16 vs 8). It is therefore not surprising that the computing time required to implement the NIGN-WW(II) for the HIP_7 data is considerably greater than for the no covariate model. This is why, as an illustration of the computations, we set the termination criteria for the M-step at 0.1, and stopped the EM algorithm after six iterations.

Even though the HIP_7 data have not been analysed properly with the NIGN-WW(II) model at this stage, we have reached the aim of this section, which is to demonstrate the application of the EM algorithm in the NIGN-WW(II) model. We can conclude that both the theory and program related to the NIGN-WW(II) do work well.

7.9 Summary

In this chapter, the PPO-survival model(s) formulated in Chapter 5, have thus been extended to PPO-survival model(s) (II) by adding in time independent pre-treatment covariates X_i. The accelerated failure time (AFT) (p.64, Cox & Oakes (1984) or p.46, Klein (1997)) has played an important role in the development. Because pre-treatment covariates X_i are involved, besides the assigned treatment $Z_i = z$ and the actual received treatment $D_i(z) = d$, in this case, the conditional probability of compliance type must be conditional on the pre-treatment covariates X_i. This leads to a much more complicated formulation than that discussed in Chapter 5.

Importantly, we assume that the conditional probability of compliance type, given $Z_i = z$, $D_i(z) = d$ and $X_i = \mathbf{x}$, which is denoted by $\pi_{c,zdx}$, follows a logistic model as in Mealli et al. (2004). This is justisfied because given $Z_i = z$, $D_i(z) = d$, compliance type has up to two categories, and $\pi_{c,11x}$ and $\pi_{c,00x}$ both satisfy a logistic model.

Our PPO-survival model(s) (II) assume a Weibull and a log-normal distribution for the failure T_{0i} and the censoring time C_{0i} (corresponding to the baseline), and also assume T_{0i} and C_{0i} to satisfy the accerlerated failure time (AFT) model. By following Theorem 1 (Section 5.7.1) and the conclusion given in Chapter 5, which states that when T_{0i} and C_{0i} follow a Weibull (or a log-normal) distribution, the failure and censoring time corresponding to the observed subgroups all follow a Weibull (or a log-normal) distribution, we can therefore conclude that the failure and censoring time corresponding to the observed subgroups with pre-treatment covariates X_i also follow the same distribution types. These results allow us to formally create the link between the PPO-survival model with covariates and those without covariates (the PPO-survival model(s) (I)). These equivalences facilitate specification of the likelihood. Our initial exploration of the application of the EM algorithm to the extended model suggests that the EM algorithm is a feasible approach to model-fitting, although computation times will clearly be greater than in the no-covariate case.

Chapter 8

Discussion and Conclusions

8.1 Overview and Summary

This thesis has used the framework of potential outcomes to develop parametric survival models for modelling treatment effects in RCTs with time-to-event outcomes which allow for noncompliance and non-ignorable censoring. As in most of the literature on the analysis of RCTs (Mealli et al., 2004; Peng et al., 2004; Barnard et al., 2003; Loeys & Goetghebeur, 2003; Yau & Little, 2001), the modelling approach developed in this thesis deals only with all-or-none compliance. Extensions to handle partial compliance are left for future work. While challenging, these extensions should follow the general framework outlined here with the concept of *compliance type* (Section 1.7.2) being extended to accommodate the quantitative notion of "degree of compliance."

The fundamental idea underpinning this thesis and previous work on causal inference, under noncompliance and subsequent missing outcomes (Frangakis & Rubin, 1999; Barnard et al., 2003; O'Malley & Normand, 2005), is that adjustment for bias related to post-randomisation self-selection of treatment, can be achieved by the introduction of a compliance type covariate, which characterises treatment received, under all possible treatment assignments. Censoring is assumed to be ignorable conditional on compliance type *latent ignorability*, but because this covariate is only partially observed, the assumed censoring mechanism is, in general, not ignorable given the observable data. This leads to a more complex likelihood construction than in standard time-to-event analyses because the model for the censoring time does not leave out of the likelihood. As a consequence, the censoring distribution needs to be explicitly modelled (Section 5.5). While the assumption

of *latent ignorability* cannot be verified from observed data, we agree with Frangakis & Rubin (1999) that because compliance type may be associated with loss to follow-up, the ignorability of the censoring mechanism conditional on compliance type will generally be a more plausible assumption than unconditional ignorability. We note, however, that in some randomised time-to-event studies with noncompliance, it will be reasonable to ignore the censoring mechanism. For example, in some studies, particularly those with short follow-up time, it is possible that censoring is caused solely by the conclusion of the study, in which case the censoring time is effectively a known constant for each individual (end of study date minus entry date). Therefore in this scenario, censoring time does not enter the analysis (Cox & Oakes, 1984; Klein, 1997) (Section 4.3.1).

Since the compliance type covariate is partially missing, this complicates model-fitting. Computational methods from the missing data literature (Mattei & Mealli, 2007; O'Malley & Normand, 2005; Mealli et al., 2004; Frangakis & Rubin, 2002, 1999) are then required to fit the model. In Chapter 5, we outlined an EM algorithm (Dempster et al., 1977) for fitting the model to the data with missingness. However, because of a drawback in the EM algorithm, which is that it does not produce estimates of the variances for model parameters, Louis's method (Louis, 1982) need to be adapted for our approach. We estimated the variances for the interest part of model parameters θ through the observed information matrix (OIM). Therefore the standard error corresponding to these estimated parameters are obtained straightforwardly after having their variances values. The standard error for the rest part of model parameters such as π_c need an alternative method because π_c is not involved in the EM algorithm. In this thesis we adopted the Bootstrap approach to figure out the standard error of parameter π_c.

The modelling approach developed in Chapter 5 and applied to the HIP study (in Chapter 6) is extended to deal with covariates in Chapter 7. This is an important extension as it allows for the analysis of variation in treatment effects by covariate level. Our extension to accommodate covariates takes advantage of an Accelerated Failure Time (AFT) (p.64, Cox & Oakes (1984) or p.46, Klein (1997)) representation of the basic model developed in Chapter 5. Under this representation of the model, extension to include covariates is

quite natural and parallels the approach outlined by Peng et al. (2004) for normally distributed outcomes.

In addition to developing a new modelling approach for randomised time to event studies with noncompliance, a further contribution of this thesis is to re-formulate the definition of causal effects in a time-dependent manner appropriate for time -to-event studies (Chapter 3). In the potential outcomes literature, causal effects are conventionally defined as contrasts between average responses (Rubin, 2005; Imbens & Rubin, 1997; Holland, 1986). However, in time-to-event studies, interest often centres on other aspects of outcome distributions, such as the survival or hazard function. Accordingly, we defined the Average Causal Effect at time t, ACE(t), to be the difference between the probability of surviving to time t, if assigned to the experimental treatment and the probability of surviving to time t, if assigned to the control treatment (Section 3.3.3). The Complier Average Causal Effect at time t, CACE(t), is then defined as the same contrast but restricted to the subgroup of *compliers*, (the subgroup of individuals who would always comply with their assigned treatment) (Section 3.3.4). Because assigned and received treatments are identical for compliers, it will often be reasonable to attribute non-zero values of CACE(t) to the causal effect of treatment on the probability of surviving to time t. By contrast, the ACE(t) reflects only the effect of assigned treatment, rather than the effect of actually receiving the treatment. For the subgroups who would always receive the new treatment (*always-takers*) or the control treatment (*never-takers*), regardless of their assigned treatment, it is plausible to assume that assigned treatment has no impact on outcome because, for these groups, the treatment actually received is not modified by assigned treatment. This assumption is formalised as the *exclusion restriction* assumption (equation (5.5.18) or equation (5.5.17) in Chapter 5). When this assumption is true and there are no individuals who would always receive the opposite of their assigned treatment (*defiers*), ACE(t) then mixes the causal effect of treatment for *compliers*, (with the absence of effect of assignment) for *never-takers* and *always-takers*, and hence clearly ACE(t) does not represent the causal effect of treatment.

The analytical approach developed in this thesis is fully parametric and consequently inferences are potentially sensitive to distributional assump-

tions for both the failure time and censoring-time distributions. However, in the analysis of the HIP study presented in Chapter 6, inferences for ACE(t) and CACE(t) were not sensitive to distributional assumptions. Whereas inferences for the conventional causal effect measures contrasting expected failure times were sensitive to these assumptions. This can be seen from Table 6.12 (page 172). Similar conclusions can be made by comparing the results of estimated ACE and ACE(t) for the HIP_7 via these two specific non-ignorable PPO survival models (Section 6.2.2).

In general, sensitivity to distributional assumptions can be explored by fitting the model under different distributional assumptions and comparing the usual sorts of likelihood-based model selection criteria, such as AIC and BIC. Table 6.12 shows that the AIC for the non-ignorable PPO-survival Weibull model (NIGN-WW model) is less than the AIC for the non-ignorable PPO-survival log-normal model (NIGN-LL model), i.e. $2.0949E+05 < 2.8393E+05$. This implies that the non-ignorable PPO-survival Weibull model (NIGN-WW model) performed slightly better than the non-ignorable PPO-survival log-normal model (NIGN-LL model) on the analysis of the HIP_7 data.

It is noteworthy that the PPO-survival models analysis of the HIP breast cancer data (Shapiro et al., 1988; Baker, 1998) in this thesis (see Chapter 6) are more informative than those from previous analyses of the HIP data (Shapiro et al., 1988, 1982; Shapiro, 1977). Indeed, we have shown that Subgroup 1, (individuals who took the screening examination), has a higher survival probability than the subgroups where individuals did not take the screening examination. This shows a clear treatment benefit. More specially, according to the results of the analysis of the HIP_7 dataset, we can conclude that in 10,000 women, about 14.8 ($\pm$3.83) women would be saved from breast cancer-related death, five years after screening. The result from the NIGN-LL model is similar, in that $\widehat{CACE}(5) = 15.7E-04$ ($SE_{cace(5)} = 4.02E-04$) (Table 6.3 on page 139 or Table C-23 on page 264). This is equivalent to about 15.7 ($\pm$4.02) women in 10,000 being saved from breast cancer-related death five years after screening.

8.2 Limitations

Only a single application was considered in this thesis and this means that it is not possible to draw general conclusions concerning sensitivity to distributional assumptions, the performance of the EM algorithm or frequentist performance of the likelihood-based estimators. Regarding the latter point, likelihood theory guarantees consistency of point estimates, as is also the case for the moment estimates of Frangakis-Rubin (Frangakis & Rubin, 1999). In the context of normally distributed outcomes, the recent simulation study of O'Malley & Normand (2005) suggested likelihood-based methods had practical, finite sample advantages compared to moment-based methods in terms of efficiency, mean-squared error and confidence interval coverage. Encouragingly for likelihood-based methods, these advantages persisted even when the distributional assumptions of the likelihood construction were incorrect (O'Malley & Normand, 2005). However, it is not known whether this model-robustness carries over to time-to-event data. In the analysis reported in Chapter 6, standard errors for causal effect estimates were markedly reduced under the parametric models, compared to the non-parametric Frangakis-Rubin estimates; whereas point estimates were similar, at least up to seven years of follow-up (Table 6.3). Thus it seems reasonable to conjecture that the parametric approach leads to a genuine efficiency over the non-parametric Frangakis-Rubin method (Frangakis & Rubin, 1999).

A limitation of the models proposed in this thesis is that the comparison with the nonparametric Frangakis-Rubin (F-R) method (Frangakis & Rubin, 1999) suggests that the standard parametric forms used in the HIP analysis are insufficiently flexible to model long-term follow-up data (Figure 6.21 (on page 167) and Figure 6.23 (on page 169)). It appears that the parametric and non-parametric approaches give similar results when follow-up is restricted to seven years, but results diverge when follow-up is extended to 18 years (Figures 6.21 and 6.23). The non-parametric analysis suggests the beneficial effect of mammographic screening levels off after 6 to 7 years. For both for the Weibull and log-normal models developed here, our results indicate that the effect of screening keeps increasing long after screening. The latter result is possibly less plausible than a levelling off in the so-called screening

effect (Baker, 1998) and reflects the parametric formulation of the survival functions used in the analysis. While this suggests the need for more flexible models, it does not invalidate the modelling approach, in general, as it relates only to specific distributional choices.

8.3 Extensions and directions for further work

8.3.1 Improvement in computation

The EM algorithm (outlined in Chapter 5) was generally slow to converge. The EM algorithm was implemented in the MATLAB computer software (www.mathworks.com/products/matlab). Computational performance could be improved by implementing the algorithm in a lower-level language such as C++.

Another plausible way of improving computation, would be to replace the EM algorithm with its extensions, such as the Supplemented EM (SEM) algorithm (Meng & Rubin, 1991; McLachlan & Krishnan, 1997) or the Expectation-Conditional Maximum (ECM) algorithm (Meng & Rubin, 1993; McLachlan & Krishnan, 1997). These algorithms can speed up convergence and thus would be very helpful when more components are involved in the model parameter vector. This is in the case of the parametric potential-outcome PPO-survival model (II).

8.3.2 Time-dependent covariates

The parametric potential-outcome (PPO) survival model with covariates developed in Chapter 7 was defined only for time-independent pre-treatment covariates. However, given that the model can be represented using the AFT model framework and that this framework extends to accommodate time-varying covariates (Cox & Oakes, 1984; Korhonen, 2000), it may be possible to further exploit the AFT model structure in dealing with more complicated RCTs, involving longitudinal designs and time-dependent covariates.

8.3.3 More flexible model forms

As noted above, standard choices of parametric failure time distributions seem insufficiently flexible for modelling longer-term effects of treatment. This suggests the need to develop more flexible parametric models. Alternatively, it may be advantageous to exploit the connection with AFT models discussed in Chapters 5 and 7, and to develop semi-parametric approaches by developing non-parametric approaches to estimating the baseline survival distribution (Robins & Tsiatis, 1991). It is also likely that Cox type models could be applied to modelling both the failure time and censoring time distributions, thereby leading to an alternative semi-parametric approach. However, a drawback to moving to Cox models is that these models are most easily interpreted in terms of hazard functions and computation of effect measures, such as ACE(t) and CACE(t), which are based on survival probabilities, may therefore be less convenient under Cox models. However, it would also be possible to develop causal effect definitions based on hazard ratios (Loeys & Goetghebeur, 2003).

8.3.4 Bayesian inference

Because of the rapid development of computer techniques during the last two decades, Bayesian approaches (Gelman et al., 2004) have shown proved advantageous, particularly to missing data problems (Imbens & Rubin, 1997; Gelman et al., 2004; Ibrahim et al., 2001; Graham, 2000; Gong et al., 2005). The Bayesian model treats the model parameters as random variables via a prior distribution, which follows from previous inference and is updated by the data. The prior distribution is selected to reflect the investigator's prior belief in the values of the parameters. In general, the prior distribution is expressed as a multivariate distribution on the parameters. Bayesian models thus allow incorporation of prior information in a natural way. This is well known as: "*Today's posterior is tomorrow's prior*" (Gelman et al., 2004). The likelihood formulations given in Chapters 5 and 7 provide the basis for moving to a Bayesian perspective by combining the likelihood with a prior distribution over the model parameters. Posterior simulation could proceed using a Gibbs sampling framework (Gelman et al., 2004; Arjas & Gasbarra,

1994), alternating between imputation of unobserved compliance types, and sampling parameters from the conditional posterior distribution, given the imputed compliance states. Both AFT and Cox type models could well be fitted within a Bayesian framework (Ibrahim et al., 2001) and therefore both fully parametric and semi-parametric Bayesian modelling should prove feasible using the general likelihood construction presented in Chapters 5 and 7. See a recent Bayesian adaptation by Gong et al. (2005).

8.3.5 Noncompliance and "Truncation-by-death"

Many randomised studies involving humans require treatment comparisons, not only to accommodate noncompliance (with the randomly assigned treatment), but treatment comparisons need to be adjusted for "truncation-by-death", which indicates the case where the input data missing due to death (Mattei & Mealli, 2007; Frangakis et al., 2007; Zhang & Rubin, 2003).

Whilst our work considered the outcome of interest as death (from breast cancer in the HIP trial), our model could just as well focus on patient states apart from death, and the truncation by death issue could possibly be accommodated for, whether by an application of principal stratification (Frangakis & Rubin, 2002) with a Bayesian parametric approach as in Mattei & Mealli (2007) to the time-to-event scenario (where death is not the outcome of interest) or by an extension of the work of Imai (Imai, 2008a) who recently generalised the results of Zhang & Rubin (2003) and Imai (2008) to randomised experiments with noncompliance by deriving the bounds on the average and quartile treatment effects for *compliers*. Mattei & Mealli (2007) have an approach which is an alternative to Imai's approach (Imai, 2008a), in that they consider point identification of treatment effects via (Bayesian) parametric modelling. None of the above work pertains to time-to-event scenarios (where death is not the outcome of interest). See also Imai (2007) and Imai (2008).

It is noteworthy that the approach proposed by Mattei & Mealli (2007) applied to the Faenza randomised experiment on Breast Self-Examination (Ferro et al. (1996) and Mealli et al. (2004)) did not include covariates, and their structural and functional assumptions (stable unit treatment value

assumption (SUTVA) formalised with potential outcomes by Rubin (1978)) alone do not allow for full identification of the underlying parameters of interest.

Recently, Frangakis et al. (2007) proposed a framework for addressing the issue of missingness caused by death. Their approach is to obtain and use data and explicit assumptions bout a treatment mechanism that could cause missing values to different levels of the treatment had been assigned. We could address our PPO-survival models to account for so-called non-death events (earlier heath states) but allow for "truncation-by-death", noncompliance and so accommodate missingness within a so-called potential outcomes framework.

8.4 Conclusions

In brief, by exploiting the potential outcomes framework and carefully considering the likelihood construction for RCTs with noncompliance in time-to-event studies, we have developed a parametric, likelihood-based approach, namely the parametric potential-outcome (PPO) survival model, for the analysis of such studies. Because censoring is common in time-to-event studies and censoring gives rise to a type of missing outcome data, the additional complications of missing outcomes cannot be ignored in RCTs with failure time outcomes. In the modelling framework developed in Chapter 5, the possibility of non-ignorable censoring is accommodated by the assumption of *latent ignorability* (Frangakis & Rubin, 1999). This model structure is more general than standard approaches for analysing time-to-event data which assume ignorable censoring (Loeys & Goetghebeur, 2003).

Noncompliance is common in RCTs and proper analysis of such studies is crucial to obtaining valid estimates of treatment effects. The models and methods developed in this thesis therefore contribute to the literature on an important and commonly occurring problem in the analysis of RCTs. The approach developed in Chapter 5 of this thesis can be viewed as a parametric alternative to the moment-estimation method proposed by Frangakis & Rubin (1999) and extends the likelihood-based approach of O'Malley & Normand (2005) to the case of failure time data. The extended model with

pre-treatment covariates derived in Chapter 7 of this thesis can be viewed as an approach which was developed from the combination of Peng et al. (2004) and O'Malley & Normand (2005) to time-to-event studies.

Although the issue of treatment noncompliance shifts the analysis of randomised studies towards the analysis of observational studies, the randomisation of assigned treatment gives randomised studies a distinct advantage over observational studies; because it ensures that the distribution of compliance type can be assumed to be the same in each treatment arm. Together with plausible assumptions, such as the *exclusion restriction* assumption (Angrist et al., 1996; Frangakis & Rubin, 1999), this ensures that the observed study data are informative about the distribution of compliance type (equation (5.5.18)).

The observational study counterpart to noncompliance in randomised studies is the problem of confounding by unobserved covariates, and because the observed study data are uninformative about unobserved covariates, causal inference in the observational study setting can be highly sensitive to prior assumptions about the adequacy of control for confounding variables (Greenland et al., 1999; Greenland, 2004). Therefore, while noncompliance complicates the analysis of RCTs, the RCTs with noncompliance are still likely to yield stronger inferences concerning treatment effects than observational studies.

Bibliography

Aalen, O. (1975). *Statistical Inference for a Family of Counting Processes.* PhD thesis, University of California.

Aalen, O. (1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics*, 6, 701–726.

Acquavella, J., Doe, J., Tomenson, J., Chester, G., Cowell, J., & Bloemen, L. (2003). Epidemiologic studies of occupational pesticide exposure and cancer: Regulatory risk assessments and biologic plausibility. *Annals of Epidemiology*, 13(1), 1–7.

Aiello, A. E. & Larson, E. L. (2002). Causal inference: The case of hygiene and health. *American Journal of Infection Control*, 30(8), 503–510.

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434), 444–455.

Arjas, E. & Gasbarra, D. (1994). Nonparametric Bayesian inference from right censored survival data, using the Gibbs sampler. *Statistica Sinica*, 4, 505–524.

Baker, S. G. (1992). A simple method for computing the observed information matrix when using the EM algorithm with categorical data. *The American Statistical Association*, 1(1), 63–67.

Baker, S. G. (1994). Gender, ethnicity, and homelessness - accounting for demographic diversity on the streets. *American Behavioral Scientist*, 37(4), 476–504.

Baker, S. G. (1997). All-or-none compliance. In S. Kotz, C. Read, & D. Banks (Eds.), *The Encyclopedia of Statistical Sciences*, volume 1 (pp. 134–138). New York: Wiley.

Baker, S. G. (1998). Analysis of survival data from a randomised trial with all-or-none compliance: Estimating the cost-effectiveness of a cancer screening program. *Journal of the American Statistical Association*, 93(443), 929–934.

Baker, S. G. & Kramer, B. S. (2005). Simple maximum likelihood estimates of efficacy in randomised trials and before-and-after studies, with implications for meta-analysis. *Statistical Methods in Medical Research*, 14, 349–367.

Barnard, J., Frangakis, C. E., Hill, J. L., & Rubin, D. B. (2003). Principal stratification approach to broken randomised experiments: A case study of School Choice vouchers in New York City. *Journal of the American Statistical Association*, 98(462), 299–311.

Breslow, N. E. (2003). Presidential Address: XXI International Biometric Conference, Freiburg, Germany, July 2002 - Are statistical contributions to medicine undervalued? *Biometrics*, 59(1), 1–8.

Casella, G. & Berger, R. L. (2002). *Statistical inference*. Australia ; Pacific Grove, CA: Thomson Learning, 2nd edition.

Chib, S. & Hamilton, B. H. (2002). Semiparametric Bayes analysis of longitudinal data treatment models. *Journal of Econometrics*, 110(1), 67–89.

Cobos, P. L., Lopez, F. J., Cano, A., Almaraz, J., & Shanks, D. R. (2002). Mechanisms of predictive and diagnostic causal induction. *Journal of Experimental Psychology-Animal Behavior Processes*, 28(4), 331–346.

Cox, D. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society*, 34, 187.

Cox, D. & Oakes, D. (1984). *Analysis of Survival Data*. Monographs on Statistics and Applied Probability. London; New York: Chapman and Hall.

Dawid, A. (2000). Causal inference without counterfactuals (with discussion). *Journal of the American Statistical Association*, 95(No.450), 407–448.

Dehejia, R. H. & Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics*, 84(1), 151–161.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via EM algorithm. *Journal of the Royal Statistical Society Series B-Methodological*, 39(1), 1–38.

Draper, D. (1995). Inference and hierarchical modeling in the social-sciences. *Journal of Educational and Behavioral Statistics*, 20(2), 115–147.

Dunn, G. & Goetghebeur, E. (2005a). Analysis compliance in clinical trials. *Statistical Methods in Medical Research*, 14, 325–326.

Dunn, G., Maracy, M., Dowrick, C., Ayuso-Mateos, J. L., Dalgard, O. S., Page, H., Lehtinen, V., Casey, P., Wilkinson, C., Vazquez-Barquero, J. L., Wilkinson, G., & Grp, O. (2003). Estimating psychological treatment effects from a randomised controlled trial with both non-compliance and loss to follow-up. *British Journal of Psychiatry*, 183, 323–331.

Dunn, G., Maracy, M., & Tomenson, B. (2005b). Estimating treatment effects from randomised clinical trials with noncompliance and loss to follow-up: the role of instrumental variable methods. *Statistical Methods in Medical Research*, 14, 369–395.

Efron, B. & Tibshirani, R. (1993). *An Introduction to the Bootstrap*. New York: Chapman& Hall.

Ferro, S., Caroli, A., Nanni, O., Biggeri, A., & Gambi, A. (1996). A randomised trial on breast self examination in Faenza (Nothern-Italy). *Tumori*, 82, 329–334.

Follmann, D. (2000). On the effect of treatment among would-be treatment compliers: An analysis of the multiple risk factor intervention trial. *Journal of the American Statistical Association*, 95(452), 1101–1109.

Frangakis, C. E. & Rubin, D. B. (1999). Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika*, 86(2), 365–379.

Frangakis, C. E. & Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics*, 58(1), 21–29.

Frangakis, C. E., Rubin, D. B., An, M. W., & Mackenzie, E. (2007). Principle stratification designs to estimate input data missing due to death. *Biometrics*, 63(Septemper), 641–662.

Fraser, D. (1957). *Nonparametric Methods in Statistics*. New York: Wiley.

Gelman, A., Carlin, J. B., Stern, H., & Rubin, D. (2004). *Bayesian Data Analysis*. London: Chapman & Hall, 2nd edition.

Geng, Z. & Jing, H. (2002). Statistical causal inference (in Chinese). In J. Fang & Y. Lu (Eds.), *Advanced Medical Statistics* (pp. 512–534). Beijing: People's Medical Publishing House.

Gong, Z., Hudson, I., & Graham, P. (2005). A Bayesian Causal Cox Model (CCM) for all-or-nothing compliance. In A. Francis, K. Matawie, A. Oshlack, & G. Smyth (Eds.), *20th International Workshop in Statistical Modelling, Sydney, Australia* (pp. 227–230).

Graham, P. (2000). Bayesian inference for a generalized population attributable fraction: the impact of early vitamin A levels on chronic lung disease in very low birth weight infants. *Statistics in Medicine*, 19, 937–956.

Graham, P. (2001). *Bayesian Causal Inference for Epidemiological and Clinical Studies*. Phd thesis, Otago, NZ.

Greenland, S. (2004). An overview of methods for causal inference from observational studies. In A. Gelman & X. L. Meng (Eds.), *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives.* John Wiley and Sons.

Greenland, S., Robins, J. M., & Pearl, J. (1999). Confounding and collapsibility in causal inference. *Statistical Science*, 14(1), 29–46.

Hamilton, B. H. (2001). Estimating treatment effects in randomised clinical trials with non-compliance: The impact of maternal smoking on birthweight. *Health Economics*, 10(5), 399–410.

Hassin, R. R., Bargh, J. A., & Uleman, J. S. (2002). Spontaneous causal inferences. *Journal of Experimental Social Psychology*, 38(5), 515–522.

Heitjan, D. F. (1999). Causal inference in a clinical trial: A comparative example. *Controlled Clinical Trials*, 20(4), 309–318.

Hernan, M., Cole, S. R., Margolick, J., Cohen, M., & Robins, J. (2005). Structural accelerated failure time models for survival analysis in studies with time-varying treatments. *Pharmacoepidemiology and Drug Safety*, Published online in Wiley Interscience (www.interscience.wiley.com), DOI:10.1002/pds.1064.

Hernan, M. A., Brumback, B., & Robins, J. M. (2000). Marginal structural models to estimate the causal effect of Zidovudine on the survival of HIV-positive men. *Epidemiology*, 11(5), 561–570.

Hernan, M. A., Brumback, B., & Robins, J. M. (2001). Marginal structural models to estimate the joint causal effect of nonrandomized treatments. *Journal of the American Statistical Association*, 96(454), 440–448.

Hernan, M. A., Hernandez-Diaz, S., Werler, M. M., Robins, J. M., & Mitchell, A. A. (2000a). Causal knowledge as a prerequisite for confounding evaluation. examples from birth defects epidemiology. *American Journal of Epidemiology*, 151(11), 43–43.

Higham, D. & Higham, N. (2000). *MATLAB Guide*. SIAM.

Hirano, K. & Imbens, G. W. (2000). Assessing the effect of an influenza vaccine in an encouragement design. *Biometrics*, 1, 69–88.

Hogan, J. W. & Daniels, M. J. (2002). A hierarchical modelling approach to analysing longitudinal data with drop-out and non-compliance with application to an equivalence trial in paediatric acquired immune deficiency syndrome. *Journal of the Royal Statistical Society. Series C, Applied Statistics*, 51, 1–21.

Holland, P. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81(396), 945–970.

Holman, C. D. J., Arnold-Reed, D. E., de Klerk, N., English, D. R., & Donovan, R. J. (2001). Epidemiologists' characteristics had little influence on causal inference. *Epidemiology*, 12(6), 752–753.

Ibrahim, J. G., Chen, M. H., & Sinha, D. (2001). *Bayesian Survival Analysis*. New York: Springer-Verlag.

Imai, K. (2007). Statistical analysis of randomised experiments with nonignorable missing binary outcomes. *Technical Report*, Department of Politics, Princeton University(available at http: ssrn.com/abstract=1014849).

Imai, K. (2008). Sharp bounds on causal effects in randomised experiments with "truncation-by-death". *Statistics and Probability Letters*, 78(2), 144–149.

Imai, K. (2008a). Identification analysis for randomised experiments with noncompliance and "truncation-by-death". Forthcoming(available at http://imai.princeton.edu/research).

Imbens, G. & Rubin, D. (1997). Bayesian inference for causal effects in randomised experiments with noncompliance. *The Annals of Statistics*, 25(1), 305–327.

Iribarren, C., Friedman, G. D., Klatsky, A. L., & Eisner, M. D. (2001). Exposure to environmental tobacco smoke: association with personal characteristics and self reported health conditions. *Journal of Epidemiology and Community Health*, 55(10), 721–728.

Jeffrey, M. (2007). Mediation analysis via potential outcomes models. *Statistics in Medicine*, Published online in Wiley Interscience (www.interscience.wiley.com), DOI:10.1002/sim.3016.

Jo, B. (2002). Statistical power in randomised intervention studies with noncompliance. *Psychological Methods*, 7(2), 178–193.

Jo, B. (2002a). Estimation of intervention effects with noncompliance: Alternative model specifications. *Journal of Educational and Behavioral Statistics*, 27(4), 385–409.

Jo, B. (2002b). Estimation of intervention effects with noncompliance: Alternative model specifications - Rejoinder by Booil Jo. *Journal of Educational and Behavioral Statistics*, 27(4), 417–420.

Joffe, M. M. (2001). Using information on realised effects to determine prospective causal effects. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 63, 759–774.

Joffe, M. M. & Brensinger, C. (2003b). Weighting in instrumental variables and G-estimation. *Statistics in Medicine*, 22(8), 1285–1303.

Joffe, M. M., Ten Have, T. R., & Brensinger, C. (2003). The compliance score as a regressor in randomised trials. *Biostatistics*, 4(3), 327–340.

Kalbfleisch, J. D. & Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. New York: Wiley, 2nd edition.

Kaufman, J. S. & Cooper, R. S. (2001). Commentary: Considerations for use of racial/ethnic classification in etiologic research. *American Journal of Epidemiology*, 154(4), 291–298.

Klein, J. P. (1997). *Survival Analysis*. Statistics for Biology and Health. New York: Springer-Verlag, 1st edition.

Korhonen, P. & Laird, N.M. Palmgren, J. (1999). Correcting for non-compliance in randomised trials: an application to the ATBC study. *Statistics in Medicine*, 18, 2879–2897.

Korhonen, P. A. (2000). *Accelerated Failure Time Models for Non-ignorable Non-compliance in Randomised Sstudies*. PhD thesis, University of Helsinki.

Kraft, C. H. & Eeden, C. v. (1968). *A Nonparametric Introduction to Statistics*. New York: Macmillan.

Lagarias, J. G., Reeds, J. A., Wright, M. H., & Wright, P. E. (1998). Convergence properties of the Nelder-Mead simplex method in low dimensions. *Society for Industrial and Applied Mathematics*, 9(1), 112–147.

Laird, N. M. & Louis, T. A. (1982). Approximate posterior distributions for incomplete data. *Journal of Royal Statistical Society , Series B*, 44(2), 190–200.

Levy, D. E., O'Malley, A. J., & Normand, S. T. (2004). Covariate adjustment in clinical trials with non-ignorable missing data and non-compliance. *Statistics in Medicine*, 23, 2319–2339.

Li, C. Y. & Sung, F. C. (2003). Association between occupational exposure to power frequency electromagnetic fields and amyotrophic lateral sclerosis: A review. *American Journal of Industrial Medicine*, 43(2), 212–220.

Little, R. J. & Rubin, D. B. (2000). Causal effects in clinical and epidemiological studies via potential outcomes: Concepts and analytical approaches. *Annual Review of Public Health*, 21, 121–145.

Little, R. J. A. & Rubin, D. B. (2002). *Statistical analysis with missing data*. New York: Wiley, 2nd edition.

Loeys, T. & Goetghebeur, E. (2003). A causal proportional hazards estimator for the effect of treatment actually received in a randomised trial with all-or-nothing compliance. *Biometrics*, 59(1), 100–105.

Loeys, T., Vansteelandt, S., & Goetghebeur, E. (2001). Accounting for correlation and compliance in cluster randomised trials. *Statistics in Medicine*, 20(24), 3753–3767.

Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B.*, 44(2), 226–233.

Mark, S. D. & Robins, J. M. (1993). A method for the analysis of randomized trials with compliance information - an application to the multiple risk factor intervention trial. *Controlled Clinical Trials*, 14(2), 79–97.

Mattei, A. & Mealli, F. (2007). Application of principal stratification approach to the Faenza randomised experiment on breast self-examination. *Biometrics*, 63(June), 437–446.

McLachlan, G. J. & Krishnan, T. (1997). *The EM Algorithm and Extensions*. New York: Wiley.

Mealli, F., Imbens, G. W., Ferro, S., & Biggeri, A. (2004). Analyzing a randomised trial on breast self-examination with noncompliance and missing outcomes. *Biostatistics*, 5(2), 207–222.

Meng, X.-L. & Rubin, D. B. (1991). Using EM to obtain asymptotic variance-covariance matrices: the SEM algorithm. *Journal of the American Statistical Association*, 86(December), 899–909.

Meng, X.-L. & Rubin, D. B. (1993). Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika*, 80(2), 267–278.

O'Malley, A. J. & Normand, S. T. (2005). Likelihood methods for treatment noncompliance and subsequent nonresponse in randomised trials. *Biometrics*, 61(June), 325–334.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, 82, 669–710.

Pearl, J. (2000). *Causality: Models, Reasoning and Inference*. Cambridge University Press.

Pearl, J. (2001). Causal inference in health sciences: a conceptual introduction. *Health Services and Outcomes Methodology*, 2, 189–220.

Peng, Y., Little, R. J. A., & Raghunathan, T. E. (2004). An extended general location model for causal inferences from data subject to noncompliance and missing values. *Biometrics*, 60(September), 598–670.

Peng, Y., Su, S. H., & Zhang, C. F. (2004a). The relationships between college students' attributional style, self-efficacy and subjective well-being. *International Journal of Psychology*, 39(5-6).

Persson, T. & Tabellini, G. (2002). Do constitutions cause large governments? Quasi-experimental evidence. *European Economic Review*, 46(4-5), 908–918.

Robins, J. (1998). Structural nested failure time models. In P. Anderson & N. Keiding (Eds.), *Survival Analysis*, The Encyclopedia of Biostatistics (pp. 4372–4389). Chichester, UK: John Wiley and Sons.

Robins, J. (1998a). Testing and estimation of directs by reparameterizing directed acyclic graphs with structural nested models. In C. Glymour & G. Cooper (Eds.), *Computation, Causation, and Discovery*. Cambrige, MA.: The MIT Press.

Robins, J. & Rotnitzky, A. (2004). Estimation of treatment effects in randomised trials with non-compliance and a dichotomous outcome using structural mean models. *Biometrika*, 91(4), 763–783.

Robins, J. & Tsiatis, A. A. (1992). Semiparametric estimation of an accelerted failure time model with time-dependent covariates. *Biometrika*, 79, 311–319.

Robins, J. M. (1997). Non-response models for the analysis of non-monotone non-ignorable missing data. *Statistics in Medicine*, 16(1-3), 21–37.

Robins, J. M. (1999). Marginal structural models versus structural nested models as tools for causal inference. In M. Halloran & D. Berry (Eds.), *The Enviroment and Clinical Trials*. New York: Springer-Verlag.

Robins, J. M. (1999a). Association, causation, and marginal structural models. *Synthese*, 121(1-2), 151–179.

Robins, J. M. & Finkelstein, D. M. (2000). Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics*, 56(3), 779–788.

Robins, J. M., Greenland, S., & Hu, F. C. (1999b). Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome. *Journal of the American Statistical Association*, 94(447), 687–700.

Robins, J. M. & Tsiatis, A. A. (1991). Correcting for noncompliance in randomised trials using rank preserving structural failure time models. *Communications in Statistics-Theory and Methods*, 20(8), 2609–2631.

Rotnitzky, A., Holcroft, C. A., & Robins, J. M. (1997). Efficiency comparisons in multivariate multiple regression with missing outcomes. *Journal of Multivariate Analysis*, 61(1), 102–128.

Rubin, D. (1974). Estimating causal effects of treatments in randomised and nonrandomized studies. *Journal of Educational Psychology*, 66(5), 688–701.

Rubin, D. (1978). Bayesian inference for causal effects: the role of randomisation. *The Annals of Statistics*, 6(1), 34–58.

Rubin, D. (1986). Which ifs have causal answers? *Journal of the American Statistical Association*, 81, 961–962.

Rubin, D. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, 100(469), 322–331.

Rubin, D. B. (1998). More powerful randomisation-based p-values in double-blind trials with non-compliance. *Statistics in Medicine*, 17, 371–385.

Rubin, D. B. (2004). Direct and indirect causal effect via potential outcomes. *The Scandinavian Journal of Statistics*, 31, 161–170.

Scharfstein, D., Robins, J. M., Eddings, W., & Rotnitzky, A. (2001). Inference in randomised studies with informative censoring and discrete time-to-event endpoints. *Biometrics*, 57(2), 404–413.

Schochat, T. & Seidel, H. J. (2001). Epidemiology in the study of human ecology - concept and evaluation of a seminar. *Gesundheitswesen*, 63(8-9), 561–566.

Schuck, A. M. & Widom, C. S. (2001). Childhood victimization and alcohol symptoms in females: causal inferences and hypothesized mediators. *Child Abuse & Neglect*, 25(8), 1069–1092.

Shapiro, S. (1977). Evidence on screening for breast cancer from a randomized trial. *Cancer*, 39(June Supplement), 2772–2782.

Shapiro, S., Venet, W., Strax, P., & Venet, L. (1988). *Periodic Screening for Breast Cancer: The Health Insurance Plan Project and Its Sequelae, 1963-1986*. The Johns Hopkins Series in Contemporary Medicine and Public Health. Baltimore: The Johns Hopkins University Press.

Shapiro, S., Venet, W., Strax, P., Venet, L., & Roeser, R. (1982). Ten to fourteen year effect of screening on breast cancer mortality. *Journal of the National Cancer Institute*, 69(No.2), 349–355.

Shrout, P. E. (1998). Causal modeling of epidemiological data on psychiatric disorders. *Social Psychiatry and Psychiatric Epidemiology*, 33(8), 400–404.

Sommer, A. & Zeger, S. (1991). On estimating efficacy from clinical trials. *Statistics in Medicine*, 10, 45–52.

Suter, G. W., Norton, S. B., & Cormier, S. M. (2002). A methodology for inferring the causes of observed impairments in aquatic ecosystems. *Environmental Toxicology and Chemistry*, 21(6), 1101–1111.

Tanner, M. A. (1993). *Tools for statistical inference : methods for the exploration of posterior distributions and likelihood functions*. Springer series in statistics. New York: Springer-Verlag, 2nd edition.

Terracini, B. & Zanetti, R. (2003). A short history of pathology registries, with emphasis on cancer registries. *Sozial-und Praventivmedizin*, 48(1), 3–10.

Vandebosch, A., Goetghebeur, E., & Damme, L. V. (2005). Structural accelerated failure time models for the effects of observed exposures on repeated events in a clinical trial. *Statistics in Medicine*, 24, 1029–1046.

Waitzkin, H., Iriart, C., Estrada, A., & Lamadrid, S. (2001). Social medicine then and now: Lessons from Latin America. *American Journal of Public Health*, 91(10), 1592–1601.

Weed, D. L. (2002). Environmental epidemiology - basics and proof of cause-effect. *Toxicology*, 181, 399–403.

White, I. (2005). Uses and limitations of randomisation-based efficacy estimators. *Statistical Methods in Medical Research*, 14, 327–347.

White, P. A. (2001). Causal judgements about relations between multilevel variables. *Journal of Experimental Psychology-Learning Memory and Cognition*, 27(2), 499–511.

Yau, L. H. Y. & Little, R. J. (2001). Inference for the complier-average causal effect from longitudinal data subject to noncompliance and missing data, with application to a job training assessment for the unemployed. *Journal of the American Statistical Association*, 96(456), 1232–1244.

Zhang, J. & Rubin, D. B. (2003). Estimation of causal effects via principal stratification when some outcomes are truncated by "death". *Journal of Educational and Behavioral Statistics*, 28(4), 353–368.

APPENDICES

Appendix A

A Correction of equation (3.2') in Louis (1982)

The likelihood function can be expressed as the product of each individual's likelihood function $L^F(\beta) = \prod_{i=1}^{n} L_i^F(\beta)$. After taking the logarithm of the complete data likelihood $L^F(\beta)$, the observed information matrix (OIM), I_{obs}, (given in equation (5.9.17) in Section 5.9.4) can be re-written as below:

$$I_{obs} = I(\hat{\beta}) = I_1 - I_2$$

where

$$I_1 = E\left\{\left(-\sum_{i=1}^{n}\frac{\partial^2}{\partial\beta^2} logL_i^F(\beta)\right)|u \in \{c,\, a,\, n\}, \beta^k\right\},$$
$$I_2 = E\left\{\left(\sum_{i=1}^{n}\frac{\partial}{\partial\beta} logL_i^F(\beta)\right)^T\left(\sum_{j=1}^{n}\frac{\partial}{\partial\beta} logL_j^F(\beta)\right)|u \in \{c,\, a,\, n\}, \beta^k\right\}. \tag{A.1}$$

Note that the descriptive index with respect to the summation in the second term in equation (A.1) above are not the same, (denoted by i and j) here. By dividing the so-called index set, 1) $i = j$; 2) $i \neq j$, we can re-write the term, $I_2 = E_1 + E_2$, which is equivalent to:

$$E\left\{\left(\sum_{i=1}^{n}\frac{\partial}{\partial\beta} logL_i^F(\beta)\right)^T\left(\sum_{j=1}^{n}\frac{\partial}{\partial\beta} logL_j^F(\beta)\right)|u \in \{a,\, c,\, n\}, \beta^k\right\}$$
$$= E_1 + E_2$$

where

$$E_1 = \sum_i E\left\{\left(\frac{\partial}{\partial\beta} logL_i^F(\beta)\right)^T\left(\frac{\partial}{\partial\beta} logL_i^F(\beta)\right)|u \in \{a,\, c,\, n\}, \beta^k\right\}$$
$$E_2 = \sum_{i\neq j}\left\{E\left(\frac{\partial}{\partial\beta} logL_i^F(\beta)\right)^T|u \in \{a,\, c,\, n\}, \beta^k\right\}\left\{E\left(\frac{\partial}{\partial\beta} logL_j^F(\beta)\right)|u \in \{a,\, c,\, n\}, \beta^k\right\}. \tag{A.2}$$

Note that in the first term in equation (A.2), the summation is overall observations (for $i = 1$ to n); whereas in the second term, it is only a summation

for $i \neq j$.

Suppose that the parameter β is a vector with p dimensions. Then $\frac{\partial}{\partial\beta} logL_i(\beta)$ denotes a row vector with p components, for different individuals, say the i^{th} and j^{th}, in general, $\frac{\partial}{\partial\beta} logL_i^F(\beta) \neq \frac{\partial}{\partial\beta} logL_j^F(\beta)$. Therefore, their corresponding expected vectors are not the same, that is:

$$E\left\{\frac{\partial}{\partial\beta} logL_i^F(\beta)|u \in \{c,\, n,\, a\}, \beta^k\right\} \neq E\left\{\frac{\partial}{\partial\beta} logL_j^F(\beta)|u \in \{c,\, n,\, a\}, \beta^k\right\}.$$

Consequently, it is easy to produce:

$$\begin{aligned} &E\left\{\left(\frac{\partial}{\partial\beta} logL_i^F(\beta)\right)^T |u \in \{c,\, n,\, a\}, \beta^k\right\} E\left\{\left(\frac{\partial}{\partial\beta} logL_j^F(\beta)\right) |u \in \{c,\, n,\, a\}, \beta^k\right\} \\ &\neq E\left\{\left(\frac{\partial}{\partial\beta} logL_j^F(\beta)\right)^T |u \in \{c,\, n,\, a\}, \beta^k\right\} E\left\{\left(\frac{\partial}{\partial\beta} logL_i^F(\beta)\right) |u \in \{c,\, n,\, a\}, \beta^k\right\}, \end{aligned} \tag{A.3}$$

which implies that the product of $E\left\{\left(\frac{\partial}{\partial\beta} logL_i^F(\beta)\right)^T |u \in \{c,\, n,\, a\}, \beta^k\right\}$ and $E\left\{\left(\frac{\partial}{\partial\beta} logL_j^F(\beta)\right) |u \in \{c,\, n,\, a\}, \beta^k\right\}$ are not exchangeable.

Hence, we give the corrected form of equation (3.2') in Louis (1982) as follows:

$$I_{obs} = I(\hat{\beta}) = I_1 - E_1 - E_2$$

where

$$\begin{aligned} I_1 &= \sum_{i=1}^{n} E\left\{-\left(\frac{\partial^2}{\partial\beta^2} logL_i(\beta)\right) |u \in \{c,\, n,\, a\}, \beta^k\right\} \\ E_1 &= \sum_{i} E\left\{\left(\frac{\partial}{\partial\beta} logL_i^F(\beta)\right)^T \left(\frac{\partial}{\partial\beta} logL_i^F(\beta)\right) |u \in \{c,\, n,\, a\}, \beta^k\right\} \\ E_2 &= \sum_{i\neq j} \left\{E\left(\frac{\partial}{\partial\beta} logL_i^F(\beta)\right)^T |u \in \{c,\, n,\, a\}, \beta^k\right\} \left\{E\left(\frac{\partial}{\partial\beta} logL_j^F(\beta)\right) |u \in \{c,\, n,\, a\}, \beta^k\right\}. \end{aligned} \tag{A.4}$$

Note that equation (A.4) guarantees that the OIM, I_{obs}, is a $p \times p$ symmetric, but Louis's equation (3.2') in Louis (1982) cannot yield a symmetric matrix even though the author had thought I_{obs} should be symmetric (Louis, 1982). The distinction between these two equations, namely Louis's equation (3.2') in Louis (1982) and its corrected form equation (A.4) given above, are that

the last term in the former, given as:

$$2\sum_{i<j}\left\{E\left(\frac{\partial}{\partial\beta}logL_i^F(\beta)\right)^T|u\in\{c,\,n,\,a\},\beta^k\right\}\left\{E\left(\frac{\partial}{\partial\beta}logL_j^F(\beta)\right)|u\in\{c,\,n,\,a\},\beta^k\right\},$$

is not symmetric as the summation is only for $i < j$ rather than overall observations as in equation (A.4), where the last term is a symmetric matrix. This is because that after taking summation with respect to all observations both for i and j but excluding $i = j$,

$$\sum_{i\neq j}\left\{E\left(\frac{\partial}{\partial\beta}logL_i^F(\beta)\right)^T|u\in\{c,\,n,\,a\},\beta^k\right\}\left\{E\left(\frac{\partial}{\partial\beta}logL_j^F(\beta)\right)|u\in\{c,\,n,\,a\},\beta^k\right\}$$

is a symmetric matrix, even though matrix $E_\beta\left\{\frac{\partial}{\partial\beta}[logL_i^F(\beta)]\right\}^T E_\beta\left\{\frac{\partial}{\partial\beta}[logL_j^F(\beta)]\right\}$ is not symmetric for $i \neq j$.

In other words, as $E_\beta\left\{\frac{\partial}{\partial\beta}[logL_i^F(\beta)]\right\}$ is not a scalar, we cannot follow the rule of $ab + ba = 2ab$. This is the key error in equation (3.2') of Louis (1982).

Appendix B

B Iterations of the EM algorithm for the NIGN-WW and NIGN-LL models

Iteration details for the NIGN-WW model and the NIGN-LL model are given in Tables C-8 and C-9 of Appendix C. The tolerance of convergence is set at 0.005. From Tables C-8 and C-9 we see that, after 14 iterations Δ reached 0.0044 (< 0.005) for the NIGN-WW model, and one iteration leads to Δ reaching 0.0008 (< 0.005) for the NIGN-LL model. Thereby we know convergence has been achieved.

Note that the formula used to obtain Table C-8 follows the formulation of the Weibull distribution provided by the MATLAB (7.4.0, R2007a) (www.mathworks.com/products/matlab) software. This is slightly different to the form we used in equations (5.6.1), (5.6.3) and (5.6.2). MATLAB (see the MATLAB help documentation) gives the following form for the probability density function of a Weibull distribution:

$$f_w(t) = \frac{B}{A}(\frac{t}{A})^{(B-1)}exp[-(\frac{t}{A})^B]. \tag{B.1}$$

Consequently, the survival function and hazards rate function are expressed as:

$$S_w(t) = exp[-(\frac{t}{A})^B], \tag{B.2}$$

$$h_w(t) = \frac{B}{A}(\frac{t}{A})^{(B-1)}. \tag{B.3}$$

where A and B indicate the scale and shape parameters, respectively, of a Weibull distribution. The scale parameter in the two forms of the Weibull distribution, denoted by ρ (for the formulation given in Cox & Oakes (1984) used in this thesis) and A (for the form given in the MATLAB (7.4.0, R2007a)) have the following relationship, that is $A = 1/\rho$. Therefore the estimated parameters $\theta^{(1)}_{T,c1}$, $\theta^{(1)}_{T,c0}$, $\theta^{(1)}_{T,n}$ (given in Table C-8) are equivalent to the reciprocal values of $\rho_{T,c1}$, $\rho_{T,c0}$, $\rho_{T,n}$ (given in Table C-1). Similarly, the estimated parameters associated with the censoring time distribution, $\theta^{(1)}_{C,c1}$, $\theta^{(1)}_{C,c0}$, $\theta^{(1)}_{C,n}$ (given in Table C-8) are equivalent to the reciprocal values of $\gamma_{C,c1}$, $\gamma_{C,c0}$, $\gamma_{C,n}$

(given in Table C-1).

The iteration details of the EM algorithm are given in Table C-10 for the NIGN-WW model and given in Table C-11 for the NIGN-LL model. The Δ values given in the last column of Tables C-10 and C-11 show the maximum component of the difference between the two estimated parameters. The tolerance of convergence is preset at 0.005. While $Mflag = 1$ indicates that the MLEs have been achieved at each iteration.

Because of the difference of the formulation of the Weibull distribution functions provided in MATLAB, the estimated parameters shown in Table C-3 are the reciprocal of the parameters given in the last row of Table C-10.

Appendix C

C Tables

Table C-1: Estimated model parameter $\hat{\theta}$ in the NIGN-WW model for the HIP_{18} data.

Method:	NIGN-WW		Data:	HIP_{18}
$\hat{\theta}$	Estimates	SE	LCL	UCL
$\rho_{T,c1}$	0.0057	0.0005	0.0047	0.0067
$\rho_{T,c0}$	0.0064	0.0006	0.0052	0.0076
$\rho_{T,n}$	0.0064	0.0006	0.0053	0.0075
α_T	1.8254	0.0613	1.7053	1.9456
$\rho_{C,c1}$	0.0622	0.0001	0.0620	0.0625
$\rho_{C,c0}$	0.0639	0.0002	0.0636	0.0642
$\rho_{C,n}$	0.0663	0.0003	0.0657	0.0669
α_C	3.1928	0.0006	3.1916	3.1940

HIP_{18}: HIP data with follow-up of 18 years.

Table C-2: Estimated model parameter $\hat{\theta}$ in the NIGN-LL model for the HIP_{18} data.

Method:	NIGN-LL		Data:	HIP_{18}
$\hat{\theta}$	Estimates	SE	LCL	UCL
$\mu_{T,c1}$	6.2597	0.1157	6.0331	6.4864
$\mu_{T,c0}$	6.1094	0.1134	5.8871	6.3317
$\mu_{T,n}$	6.1481	0.1233	5.9065	6.3898
σ_T	1.5595	0.0479	1.4657	1.6533
$\mu_{C,c1}$	2.6173	0.0040	2.6094	2.6252
$\mu_{C,c0}$	2.6217	0.0041	2.6136	2.6298
$\mu_{C,n}$	2.3269	0.0055	2.3162	2.3376
σ_C	0.5674	0.0000	0.5673	0.5674

HIP_{18}: HIP data with follow-up of 18 years.

Table C-3: Estimated model parameter $\hat{\theta}$ in the NIGN-WW model for the HIP_7 data.

Method:	NIGN-WW		Data:	HIP_7
$\hat{\theta}$	Estimates	SE	LCL	UCL
$\rho_{T,c1}$	0.0059	0.0014	0.0031	0.0087
$\rho_{T,c0}$	0.0092	0.0020	0.0053	0.0131
$\rho_{T,n}$	0.0069	0.0016	0.0037	0.0101
α_T	1.9362	0.1389	1.6640	2.2084
$\rho_{C,c1}$	0.1438	0.0001	0.1436	0.1440
$\rho_{C,c0}$	0.1446	0.0001	0.1444	0.1448
$\rho_{C,n}$	0.1452	0.0001	0.1450	0.1454
α_C	10.3780	0.0004	10.3770	10.3790

HIP_7: HIP data with follow-up of 7 years.

Table C-4: Estimated model parameter $\hat{\theta}$ in the NIGN-LL model for the HIP_7 data.

Method:	NIGN-LL		Data:	HIP_7
$\hat{\theta}$	Estimates	SE	LCL	UCL
$\mu_{T,c1}$	6.7587	0.3394	6.0935	7.4239
$\mu_{T,c0}$	6.2969	0.3073	5.6947	6.8992
$\mu_{T,n}$	6.6068	0.3393	5.9419	7.2718
σ_T	1.6828	0.1134	1.4605	1.9052
$\mu_{C,c1}$	1.8900	0.0020	1.8861	1.8939
$\mu_{C,c0}$	1.8808	0.0017	1.8775	1.8840
$\mu_{C,n}$	1.7764	0.0034	1.7697	1.7831
σ_C	0.2831	0.0000	0.2830	0.2832

HIP_7: HIP data with follow-up of 7 years.

Table C-5: Estimated model parameter $\hat{\theta}$ in the IGN-W model for the HIP_7 data.

Method:	NIGN-W		Data:	HIP_7
$\hat{\theta}$	Estimates	SE	LCL	UCL
$\rho_{T,c1}$	0.0059	0.0011	0.0038	0.0080
$\rho_{T,c0}$	0.0098	0.0014	0.0070	0.0126
$\rho_{T,n}$	0.0039	0.0007	0.0025	0.0053
α_T	1.9326	0.0974	1.7418	2.1234

HIP_7: HIP data with follow-up of 7 years.

Table C-6: Estimated model parameter $\hat{\theta}$ in the IGN-W model for the HIP_{18} data.

Method:	NIGN-W		Data:	HIP_{18}
$\hat{\theta}$	Estimates	SE	LCL	UCL
$\rho_{T,c1}$	0.0057	0.0004	0.0049	0.0065
$\rho_{T,c0}$	0.0073	0.0005	0.0064	0.0082
$\rho_{T,n}$	0.0043	0.0003	0.0037	0.0049
α_T	1.8186	0.0491	1.7224	1.9148

HIP_{18}: HIP data with follow-up of 18 years.

Table C-7: Estimated parameters $\hat{\pi}_c$ for the HIP data

$\hat{\pi_c}$	Estimates	SE	LCL	UCL
HIP_{18}	0.6686	0.0005	0.6677	0.6696
HIP_7	0.6698	0.0000	0.6697	0.6698

HIP_{18}: HIP data with following-up 18 years.
HIP_7: HIP data with follow-up of 7 years
The SE given in the table is obtained by *bootstrap*.

Table C-8: Iterations of the EM algorithm using the NIGN-WW model for the HIP_{18} data.

Model:	NIGN-WW						Data:	HIP_{18}		
Iter k	$\theta^{(1)}_{T,c1}$	$\theta^{(1)}_{T,c0}$	$\theta^{(1)}_{T,n0}$	α_T	$\theta^{(1)}_{C,c1}$	$\theta^{(1)}_{C,c0}$	$\theta^{(1)}_{C,n0}$	α_C	Mflag [a]	Δ [b]
1	174.74	156.41	155.26	1.83	16.07	15.64	15.09	3.19	1	113.9300
2	174.75	156.20	155.52	1.83	16.07	15.64	15.09	3.19	1	0.2520
3	174.77	156.06	155.70	1.83	16.07	15.65	15.08	3.19	1	0.1870
4	174.77	155.96	155.82	1.83	16.07	15.65	15.08	3.19	1	0.1172
5	174.78	155.89	155.91	1.83	16.07	15.65	15.08	3.19	1	0.0887
6	174.78	155.85	155.96	1.83	16.07	15.65	15.08	3.19	1	0.0569
7	174.78	155.81	156.00	1.83	16.07	15.65	15.08	3.19	1	0.0390
8	174.78	155.79	156.03	1.83	16.07	15.65	15.08	3.19	1	0.0272
9	174.79	155.78	156.05	1.83	16.07	15.65	15.08	3.19	1	0.0173
10	174.78	155.77	156.06	1.83	16.07	15.65	15.08	3.19	1	0.0130
11	174.78	155.76	156.07	1.83	16.07	15.65	15.08	3.19	1	0.0094
12	174.78	155.76	156.07	1.83	16.07	15.65	15.08	3.19	1	0.0055
13	174.78	155.75	156.07	1.83	16.07	15.65	15.08	3.19	1	0.0052
14	174.78	155.75	156.08	1.83	16.07	15.65	15.08	3.19	1	0.0044

HIP_{18} : HIP data with follow-up of 18 years.

[a] $Mflag = 1$ indicates the maximum of $Q(\beta, \beta^{(k)}$ is reached in the k^{th} iteration.

[b] Δ represents the maximum component of $|\beta^{(k+1)} - \beta^{(k)}|$, the absolute vector of difference.

Table C-9: Iterations of EM algorithm for the NIGN-LL model on the HIP_{18} data

Model:	NIGN-LL						Data:	HIP_{18}		
Iter k	$\theta^{(1)}_{T,c1}$	$\theta^{(1)}_{T,c0}$	$\theta^{(1)}_{T,n0}$	σ	$\theta^{(1)}_{C,c1}$	$\theta^{(1)}_{C,c0}$	$\theta^{(1)}_{C,n0}$	σ_c	Mflag[a]	Δ[b]
1	6.2597	6.1094	6.1481	1.5595	2.6173	2.6217	2.3269	0.5674	1	0.0008

HIP_{18} : the HIP data with following-up in 18 years.

[a] $Mflag = 1$ indicates the maximum of $Q(\beta, \beta^{(k)}$ is reached in the k^{th} iteration.

[b] Δ represents the maximum component of $|\beta^{(k+1)} - \beta^{(k)}|$, the absolute vector of difference.

Table C-10: Iterations of the EM algorithm for the NIGN-WW model on the HIP_7 data

Model:	NIGN-WW						Data:	HIP_7		
Iter	$\theta^{(1)}_{T,c1}$	$\theta^{(1)}_{T,c0}$	$\theta^{(1)}_{T,n0}$	α	$\theta^{(1)}_{C,c1}$	$\theta^{(1)}_{C,c0}$	$\theta^{(1)}_{C,n0}$	α_c	Mflag[a]	Δ^b
1	169.16	109.66	142.30	1.94	6.95	6.91	6.89	10.38	1	5984.8000
2	169.18	109.32	143.14	1.94	6.95	6.91	6.89	10.38	1	0.8389
3	169.19	109.10	143.67	1.94	6.95	6.91	6.89	10.38	1	0.5288
4	169.20	108.96	144.00	1.94	6.95	6.91	6.89	10.38	1	0.3367
5	169.21	108.88	144.23	1.94	6.95	6.91	6.89	10.38	1	0.2229
6	169.22	108.83	144.36	1.94	6.95	6.91	6.89	10.38	1	0.1329
7	169.22	108.80	144.45	1.94	6.95	6.91	6.89	10.38	1	0.0887
8	169.21	108.77	144.49	1.94	6.95	6.91	6.89	10.38	1	0.0405
9	169.22	108.76	144.53	1.94	6.95	6.91	6.89	10.38	1	0.0388
10	169.21	108.75	144.54	1.94	6.95	6.91	6.89	10.38	1	0.0166
11	169.22	108.75	144.56	1.94	6.95	6.91	6.89	10.38	1	0.0190
12	169.23	108.75	144.58	1.94	6.95	6.91	6.89	10.38	1	0.0162
13	169.22	108.74	144.58	1.94	6.95	6.91	6.89	10.38	1	0.0102
14	169.22	108.74	144.58	1.94	6.95	6.91	6.89	10.38	1	0.0045

HIP_7 : HIP data with follow-up of 7 years.

[a] $Mflag = 1$ indicates the maximum of $Q(\beta, \beta^{(k)}$ is reached in the k^{th} iteration.

[b] Δ represents the maximum component of $|\beta^{(k+1)} - \beta^{(k)}|$, the absolute vector of difference.

Table C-11: Iterations of the EM algorithm for the NIGN-LL model on the HIP_7 data

Model:	NIGN-LL						Data:	HIP_7		
Iter k	$\theta^{(1)}_{T,c1}$	$\theta^{(1)}_{T,c0}$	$\theta^{(1)}_{T,n0}$	σ	$\theta^{(1)}_{C,c1}$	$\theta^{(1)}_{C,c0}$	$\theta^{(1)}_{C,n0}$	σ_c	Mflag	[a] Δ [b]
1	8.22	7.56	8.02	2.18	1.89	1.88	1.80	0.28	1	1.9602
2	6.76	6.29	6.62	1.68	1.89	1.87	1.78	0.28	1	1.4641
3	6.76	6.29	6.61	1.68	1.89	1.88	1.78	0.28	1	0.0068
4	6.76	6.30	6.61	1.68	1.89	1.88	1.78	0.28	1	0.0041

HIP_7 : HIP data with follow-up of 7 years.

[a] $Mflag = 1$ indicates the maximum of $Q(\beta, \beta^{(k)}$ is reached in the k^{th} iteration.

[b] Δ represents the maximum component of $|\beta^{(k+1)} - \beta^{(k)}|$, the absolute vector of difference.

Table C-12: Estimated ACE(t)(95% CI) using the F-R method for the HIP_{18} data.

Method:	F-R		Data:	HIP_{18}		
Year	$\widehat{ACE}_{FR}$	SE_{ace}	LCL_{ace}	UCL_{ace}	Zvalue	$p-value$
1	0.38E−04	0.93E−04	−1.44E−04	2.20E−04	0.4126	0.3399
2	2.03E−04	1.63E−04	−1.16E−04	5.23E−04	1.2464	0.1063
3	3.58E−04	2.22E−04	−0.76E−04	7.93E−04	1.6160	0.0530
4	7.38E−04	2.97E−04	1.56E−04	1.32E−03	2.4843	0.0065
5	12.1E−04	3.92E−04	4.41E−04	1.98E−03	3.0855	0.0010
6	18.6E−04	4.77E−04	9.23E−04	2.79E−03	3.8932	0.0000
7	17.1E−04	5.30E−04	6.74E−04	2.75E−03	3.2317	0.0006
8	14.1E−04	5.81E−04	2.71E−04	2.55E−03	2.4268	0.0076
9	17.3E−04	6.49E−04	4.55E−04	3.00E−03	2.6614	0.0039
10	14.6E−04	7.01E−04	0.82E−04	2.83E−03	2.0769	0.0189
11	17.3E−04	7.58E−04	2.46E−04	3.22E−03	2.2844	0.0112
12	14.9E−04	8.15E−04	−1.10E−04	3.08E−03	1.8249	0.0340
13	14.8E−04	8.68E−04	−2.19E−04	3.18E−03	1.7081	0.0438
14	10.5E−04	9.21E−04	−7.60E−04	2.85E−03	1.1350	0.1282
15	11.1E−04	10.1E−04	−8.72E−04	3.09E−03	1.0963	0.1365
16	10.5E−04	10.9E−04	−10.8E−04	3.19E−03	0.9683	0.1665
17	11.8E−04	11.4E−04	−10.6E−04	3.42E−03	1.0354	0.1502
18	21.3E−04	12.2E−04	−2.66E−04	4.52E−03	1.7426	0.0407

HIP_{18}: HIP data with follow-up of 18 years

Table C-13: Estimated ACE(t) (95% CI) using the NIGN-WW model for the HIP_{18} data.

Method:	NIGN-WW		Data:	HIP_{18}		
Year	$\widehat{ACE}_{nign-ww}$	SE_{ace}	LCL_{ace}	UCL_{ace}	Zvalue	$p-value$
1	0.13E−04	0.07E−04	−0.01E−04	0.26E−04	1.8643	0.0311
2	0.45E−04	0.24E−04	−0.02E−04	0.91E−04	1.8920	0.0292
3	0.94E−04	0.49E−04	−0.03E−04	1.91E−04	1.9030	0.0285
4	1.59E−04	0.83E−04	−0.04E−04	3.22E−04	1.9084	0.0282
5	2.38E−04	1.25E−04	−0.06E−04	4.83E−04	1.9111	0.0280
6	3.32E−04	1.74E−04	−0.08E−04	6.73E−04	1.9123	0.0279
7	4.40E−04	2.30E−04	−0.11E−04	8.91E−04	1.9127	0.0279
8	5.61E−04	2.93E−04	−0.14E−04	11.4E−04	1.9126	0.0279
9	6.95E−04	3.63E−04	−0.17E−04	14.1E−04	1.9121	0.0279
10	8.41E−04	4.40E−04	−0.21E−04	17.0E−04	1.9113	0.0280
11	10.0E−04	5.23E−04	−0.26E−04	20.3E−04	1.9104	0.0280
12	11.7E−04	6.13E−04	−0.31E−04	23.7E−04	1.9094	0.0281
13	13.5E−04	7.09E−04	−0.376E−04	27.4E−04	1.9083	0.0282
14	15.5E−04	8.11E−04	−0.43E−04	31.4E−04	1.9072	0.0282
15	17.5E−04	9.19E−04	−0.50E−04	35.5E−04	1.9061	0.0283
16	19.7E−04	10.3E−04	−0.57E−04	39.9E−04	1.9049	0.0284
17	21.9E−04	11.5E−04	−0.65E−04	44.5E−04	1.9037	0.0285
18	24.3E−04	12.8E−04	−0.73E−04	49.4E−04	1.9026	0.0285

HIP_{18}: HIP data with follow-up of 18 years

Table C-14: Estimated ACE(t) (95% CI) using the NIGN-LL model for the HIP_{18} data.

Method:	NIGN-LL		Data:	HIP_{18}		
Year	$\widehat{ACE}_{nign-ll}$	SE_{ace}	LCL_{ace}	UCL_{ace}	Zvalue	$p-value$
1	0.10E−04	0.05E−04	0.01E−04	0.19E−04	2.1060	0.0176
2	0.53E−04	0.23E−04	0.07E−04	0.98E−04	2.2561	0.0120
3	1.27E−04	0.55E−04	0.20E−04	2.34E−04	2.3214	0.0101
4	2.27E−04	0.96E−04	0.38E−04	4.16E−04	2.3576	0.0092
5	3.49E−04	1.47E−04	0.62E−04	6.37E−04	2.3804	0.0086
6	4.89E−04	2.04E−04	0.89E−04	8.89E−04	2.3958	0.0083
7	6.42E−04	2.67E−04	1.19E−04	11.7E−04	2.4068	0.0080
8	8.08E−04	3.34E−04	1.52E−04	14.6E−04	2.4150	0.0079
9	9.82E−04	4.06E−04	1.87E−04	17.8E−04	2.4213	0.0077
10	11.6E−04	4.80E−04	2.24E−04	21.1E−04	2.4262	0.0076
11	13.5E−04	5.57E−04	2.62E−04	24.5E−04	2.4302	0.0075
12	15.5E−04	6.36E−04	3.01E−04	27.9E−04	2.4334	0.0075
13	17.5E−04	7.16E−04	3.41E−04	31.5E−04	2.4361	0.0074
14	19.5E−04	7.98E−04	3.82E−04	35.1E−04	2.4384	0.0074
15	21.5E−04	8.81E−04	4.23E−04	38.8E−04	2.4404	0.0073
16	23.6E−04	9.65E−04	4.65E−04	42.5E−04	2.4421	0.0073
17	25.6E−04	10.5E−04	5.07E−04	46.2E−04	2.4435	0.0073
18	27.7E−04	11.3E−04	5.49E−04	49.9E−04	2.4448	0.0072

HIP_{18}: HIP data with follow-up of 18 years.

Table C-15: Estimated CACE(t) (95% CI) using the F-R method for the HIP_{18} data.

Method:	F-R		Data:	HIP_{18}		
Year	$\widehat{CACE}_{FR}$	SE_{cace}	LCL_{cace}	UCL_{cace}	Zvalue	$p-value$
1	0.57E−04	1.39E−04	−2.14E−04	3.29E−04	0.4127	0.3399
2	3.03E−04	2.43E−04	−1.74E−04	7.80E−04	1.2464	0.1063
3	5.35E−04	3.31E−04	−1.14E−04	11.8E−04	1.6159	0.0531
4	11.0E−04	4.44E−04	2.32E−04	19.7E−04	2.4836	0.0065
5	18.1E−04	5.85E−04	6.58E−04	29.5E−04	3.0846	0.0010
6	27.7E−04	7.13E−04	1.38E−03	41.7E−04	3.8930	0.0000
7	25.6E−04	7.91E−04	1.01E−03	41.1E−04	3.2321	0.0006
8	21.1E−04	8.68E−04	4.05E−04	38.1E−04	2.4272	0.0076
9	25.8E−04	9.69E−04	6.80E−04	44.8E−04	2.6616	0.0039
10	21.7E−04	10.5E−04	1.23E−04	42.3E−04	2.0772	0.0189
11	25.9E−04	11.3E−04	3.68E−04	48.0E−04	2.2849	0.0112
12	22.2E−04	12.2E−04	−1.64E−04	46.1E−04	1.8252	0.0340
13	22.1E−04	13.0E−04	−3.26E−04	47.5E−04	1.7082	0.0438
14	15.6E−04	13.8E−04	−11.3E−04	42.6E−04	1.1349	0.1282
15	16.5E−04	15.1E−04	−13.0E−04	46.1E−04	1.0962	0.1365
16	15.8E−04	16.3E−04	−16.1E−04	47.6E−04	0.9682	0.1665
17	17.7E−04	17.1E−04	−15.8E−04	51.1E−04	1.0353	0.1502
18	31.8E−04	18.2E−04	−3.96E−04	67.5E−04	1.7427	0.0407

HIP_{18}: HIP data with follow-up 18 years

Table C-16: Estimated CACE(t) (95% CI) using the NIGN-WW model for the HIP_{18} data.

Method:	NIGN-WW		Data:	HIP_{18}		
Year	$\widehat{CACE}_{nign-ww}$	SE_{cace}	LCL_{cace}	UCL_{cace}	Zvalue	$p-value$
1	0.19E−04	0.10E−04	−0.01E−04	0.39E−04	1.8643	0.0311
2	0.67E−04	0.35E−04	−0.02E−04	1.36E−04	1.8921	0.0292
3	1.40E−04	0.74E−04	−0.04E−04	2.85E−04	1.9031	0.0285
4	2.37E−04	1.24E−04	−0.06E−04	4.80E−04	1.9084	0.0282
5	3.56E−04	1.86E−04	−0.09E−04	7.21E−04	1.9111	0.0280
6	4.96E−04	2.59E−04	−0.12E−04	10.0E−04	1.9124	0.0279
7	6.57E−04	3.43E−04	−0.16E−04	13.3E−04	1.9128	0.0279
8	8.37E−04	4.38E−04	−0.21E−04	17.0E−04	1.9126	0.0279
9	10.4E−04	5.42E−04	−0.26E−04	21.0E−04	1.9121	0.0279
10	12.6E−04	6.57E−04	−0.32E−04	25.4E−04	1.9114	0.0280
11	14.9E−04	7.81E−04	−0.39E−04	30.2E−04	1.9105	0.0280
12	17.5E−04	9.15E−04	−0.46E−04	35.4E−04	1.9095	0.0281
13	20.2E−04	10.6E−04	−0.55E−04	40.9E−04	1.9084	0.0282
14	23.1E−04	12.1E−04	−0.64E−04	46.8E−04	1.9073	0.0282
15	26.2E−04	13.7E−04	−0.74E−04	53.0E−04	1.9061	0.0283
16	29.4E−04	15.4E−04	−0.85E−04	59.6E−04	1.9049	0.0284
17	32.8E−04	17.2E−04	−0.97E−04	66.5E−04	1.9038	0.0285
18	36.3E−04	19.1E−04	−1.09E−04	73.7E−04	1.9026	0.0285

HIP_{18}: HIP data with follow-up of 18 years.

Table C-17: Estimated CACE(t) (95% CI) using the NIGN-LL models for the HIP_{18} data.

Method:	NIGN-LL		Data:	HIP_{18}		
Year	$\widehat{CACE}_{nign-ll}$	SE_{cace}	LCL_{cace}	UCL_{cace}	Zvalue	$p-value$
1	0.15E−04	0.07E−04	0.01E−04	0.29E−04	2.1061	0.0176
2	0.78E−04	0.35E−04	0.10E−04	1.47E−04	2.2562	0.0120
3	1.89E−04	0.82E−04	0.30E−04	3.49E−04	2.3215	0.0101
4	3.39E−04	1.44E−04	0.57E−04	6.21E−04	2.3577	0.0092
5	5.22E−04	2.19E−04	0.92E−04	9.51E−04	2.3805	0.0086
6	7.30E−04	3.05E−04	1.33E−04	13.3E−04	2.3959	0.0083
7	9.59E−04	3.98E−04	1.78E−04	17.4E−04	2.4069	0.0080
8	12.1E−04	4.99E−04	2.27E−04	21.8E−04	2.4151	0.0079
9	14.7E−04	6.06E−04	2.79E−04	26.5E−04	2.4214	0.0077
10	17.4E−04	7.17E−04	3.34E−04	31.4E−04	2.4263	0.0076
11	20.2E−04	8.31E−04	3.91E−04	36.5E−04	2.4303	0.0075
12	23.1E−04	9.49E−04	4.50E−04	41.7E−04	2.4336	0.0075
13	26.1E−04	10.7E−04	5.09E−04	47.0E−04	2.4363	0.0074
14	29.1E−04	11.9E−04	5.70E−04	52.4E−04	2.4386	0.0074
15	32.1E−04	13.2E−04	6.32E−04	57.9E−04	2.4405	0.0073
16	35.2E−04	14.4E−04	6.94E−04	63.4E−04	2.4422	0.0073
17	38.3E−04	15.7E−04	7.57E−04	68.9E−04	2.4437	0.0073
18	41.4E−04	16.9E−04	8.20E−04	74.5E−04	2.4449	0.0072

HIP_{18}: HIP data with follow-up of 18 years.

Table C-18: Estimated ACE(t) (95% CI) using the F-R method for the HIP_7 data.

Method:	F-R		Data:	HIP_7		
Year	$\widehat{ACE}_{FR}$	SE_{ace}	LCL_{ace}	UCL_{ace}	Zvalue	$p-value$
1	0.38E−04	0.93E−04	−1.44E−04	2.20E−04	0.4126	0.3399
2	2.03E−04	1.63E−04	−1.16E−04	5.23E−04	1.2464	0.1063
3	3.58E−04	2.22E−04	−0.76E−04	7.93E−04	1.6160	0.0530
4	7.38E−04	2.97E−04	1.56E−04	13.2E−04	2.4843	0.0065
5	12.1E−04	3.92E−04	4.41E−04	19.8E−04	3.0855	0.0010
6	18.7E−04	4.74E−04	9.44E−04	28.0E−04	3.9507	0.0000
7	18.7E−04	4.74E−04	9.44E−04	28.0E−04	3.9507	0.0000

HIP_7: HIP data with follow-up of 7 years

Table C-19: Estimated CACE(t) (95% CI) using the F-R method for the HIP_7 data.

Method:	F-R		Data:	HIP_7		
Year	$\widehat{CACE}_{FR}$	SE_{cace}	LCL_{cace}	UCL_{cace}	Zvalue	$p-value$
1	0.57E−04	1.39E−04	−2.14E−04	3.29E−04	0.4127	0.3399
2	3.03E−04	2.43E−04	−1.74E−04	7.80E−04	1.2464	0.1063
3	5.35E−04	3.31E−04	−1.14E−04	11.8E−04	1.6159	0.0531
4	11.0E−04	4.44E−04	2.32E−04	19.7E−04	2.4836	0.0065
5	18.1E−04	5.85E−04	6.58E−04	29.5E−04	3.0846	0.0010
6	28.0E−04	7.08E−04	14.1E−04	41.9E−04	3.9505	0.0000
7	28.0E−04	7.08E−04	14.1E−04	41.9E−04	3.9505	0.0000

HIP_7: HIP data with follow-up of 7 years

Table C-20: Estimated ACE(t) (95% CI) using the NIGN-WW model for the HIP_7 data.

Method:	NIGN-WW		Data:	HIP_7		
Year	$\widehat{ACE}_{nign-ww}$	SE_{ace}	LCL_{ace}	UCL_{ace}	Zvalue	$p-value$
1	0.44E−05	0.16E−04	0.13E−04	0.75E−05	2.7777	0.0027
2	1.68E−04	0.51E−04	0.69E−04	2.67E−04	3.3248	0.0004
3	3.69E−04	1.02E−04	1.69E−04	5.68E−04	3.6237	0.0001
4	6.43E−04	1.70E−04	3.10E−04	9.76E−04	3.7839	0.0001
5	9.90E−04	2.56E−04	4.87E−04	14.9E−04	3.8593	0.0001
6	14.1E−04	3.63E−04	6.97E−04	21.2E−04	3.8822	0.0001
7	19.0E−04	4.89E−04	9.36E−04	28.5E−04	3.8729	0.0001

HIP_7: HIP data with follow-up of 7 years.

Table C-21: Estimated ACE(t) (95% CI) using the NIGN-WW model for the HIP_7 data.

Method:	NIGN-LL		Data:	HIP_7		
Year	$\widehat{ACE}_{nign-ll}$	SE_{ace}	LCL_{ace}	UCL_{ace}	Zvalue	$p-value$
1	0.41E−04	0.16E−04	0.09E−04	0.74E−04	2.5238	0.0058
2	1.86E−04	0.58E−04	0.73E−04	2.99E−04	3.2280	0.0006
3	4.15E−04	1.15E−04	1.89E−04	6.41E−04	3.5969	0.0002
4	7.07E−04	1.86E−04	3.42E−04	10.7E−04	3.7917	0.0001
5	10.5E−04	2.69E−04	5.20E−04	15.8E−04	3.8924	0.0000
6	14.3E−04	3.62E−04	7.18E−04	21.4E−04	3.9413	0.0000
7	18.4E−04	4.64E−04	9.28E−04	27.5E−04	3.9612	0.0000

HIP_7: HIP data with follow-up of 7 years

Table C-22: Estimated CACE(t) (95% CI) using the NIGN-WW model for the HIP_7 data.

Method:	NIGN-WW		Data:	HIP_7		
Year	$\widehat{CACE}_{nign-ww}$	SE_{cace}	LCL_{cace}	UCL_{cace}	Zvalue	$p-value$
1	0.66E−04	0.24E−04	0.19E−04	1.12E−04	2.7779	0.0027
2	2.51E−04	0.76E−04	1.03E−04	3.99E−04	3.3251	0.0004
3	5.50E−04	1.52E−04	2.53E−04	8.48E−04	3.6241	0.0001
4	9.60E−04	2.54E−04	4.63E−04	14.6E−04	3.7843	0.0001
5	14.8E−04	3.83E−04	7.27E−04	22.3E−04	3.8597	0.0001
6	21.0E−04	5.41E−04	10.4E−04	31.6E−04	3.8827	0.0001
7	28.3E−04	7.31E−04	14.0E−04	42.6E−04	3.8734	0.0001

HIP_7: HIP data with follow-up of 7 years

Table C-23: Estimated CACE(t) (95% CI) using the the NIGN-LL model for the HIP_7 data.

Method:	NIGN-LL		Data:	HIP_7		
Year	$\widehat{CACE}_{nign-ll}$	SE_{cace}	LCL_{cace}	UCL_{cace}	Zvalue	$p-value$
1	0.62E−04	0.25E−04	0.14E−04	1.10E−04	2.5240	0.0058
2	2.78E−04	0.86E−04	1.09E−04	4.47E−04	3.2283	0.0006
3	6.19E−04	1.72E−04	2.82E−04	9.57E−04	3.5972	0.0002
4	10.6E−04	2.78E−04	5.10E−04	16.0E−04	3.7921	0.0001
5	15.7E−04	4.02E−04	7.77E−04	23.5E−04	3.8929	0.0000
6	21.3E−04	5.41E−04	10.7E−04	31.9E−04	3.9418	0.0000
7	27.4E−04	6.92E−04	13.9E−04	41.0E−04	3.9617	0.0000

HIP_7: HIP data with follow-up of 7 years

Table C-24: Estimated CACE(t) (95% CI) using the IGN-W model for the HIP_7 data.

Method:	IGN-W		Data:	HIP_7		
Year	$\widehat{CACE}_{ign-w}$	SE_{cace}	LCL_{cace}	UCL_{cace}	Zvalue	$p-value$
1	0.82E−04	0.22E−04	0.40E−04	1.24E−04	3.7857	0.0001
2	3.13E−04	0.70E−04	1.76E−04	4.49E−04	4.4934	0.0000
3	6.85E−04	1.40E−04	4.10E−04	9.59E−04	4.8851	0.0000
4	11.9E−04	2.34E−04	7.35E−04	16.5E−04	5.1043	0.0000
5	18.4E−04	3.52E−04	11.5E−04	25.2E−04	5.2179	0.0000
6	26.1E−04	4.95E−04	16.4E−04	35.8E−04	5.2651	0.0000
7	35.1E−04	6.66E−04	22.0E−04	48.1E−04	5.2701	0.0000

HIP_7: HIP data with follow-up of 7 years.

Table C-25: Estimated CACE(t) (95% CI) using the IGN-W model for the HIP_{18} data.

Method:	IGN-W		Data:	HIP_{18}		
Year	$\widehat{CACE}_{ign-w}$	SE_{cace}	LCL_{cace}	UCL_{cace}	Zvalue	$p-value$
1	0.47E−04	0.11E−04	0.26E−04	0.68E−04	4.2754	0.0000
2	1.66E−04	0.36E−04	0.97E−04	2.36E−04	4.6724	0.0000
3	3.47E−04	0.71E−04	2.08E−04	4.86E−04	4.8988	0.0000
4	5.86E−04	1.16E−04	3.58E−04	8.13E−04	5.0498	0.0000
5	8.78E−04	1.70E−04	5.45E−04	12.1E−04	5.1582	0.0000
6	12.2E−04	2.33E−04	7.65E−04	16.8E−04	5.2392	0.0000
7	16.2E−04	3.05E−04	10.2E−04	22.1E−04	5.3014	0.0000
8	20.6E−04	3.85E−04	13.0E−04	28.1E−04	5.3501	0.0000
9	25.5E−04	4.73E−04	16.2E−04	34.7E−04	5.3885	0.0000
10	30.8E−04	5.69E−04	19.7E−04	42.0E−04	5.4191	0.0000
11	36.6E−04	6.73E−04	23.4E−04	49.8E−04	5.4436	0.0000
12	42.8E−04	7.84E−04	27.5E−04	58.2E−04	5.4631	0.0000
13	49.5E−04	9.03E−04	31.8E−04	67.2E−04	5.4786	0.0000
14	56.5E−04	10.3E−04	36.3E−04	76.7E−04	5.4909	0.0000
15	63.9E−04	11.6E−04	41.2E−04	86.7E−04	5.5004	0.0000
16	71.8E−04	13.0E−04	46.2E−04	97.3E−04	5.5078	0.0000
17	80.0E−04	14.5E−04	51.6E−04	108.0E−04	5.5132	0.0000
18	88.6E−04	16.1E−04	57.1E−04	120.0E−04	5.5171	0.0000

HIP_{18}: HIP data with follow-up of 18 years.

Table C-26: Estimated ACE(t) (95% CI) using the IGN-W model for the HIP_7 data.

Method:	IGN-W		Data:	HIP_7		
Year	$\widehat{ACE}_{ign-w}$	SE_{cace}	LCL_{cace}	UCL_{cace}	Zvalue	$p-value$
1	0.55E−04	0.15E−04	0.27E−04	0.83E−04	3.7852	0.0001
2	2.10E−04	0.47E−04	1.18E−04	3.01E−04	4.4927	0.0000
3	4.59E−04	0.94E−04	2.75E−04	6.43E−04	4.8842	0.0000
4	7.99E−04	1.57E−04	4.92E−04	11.1E−04	5.1032	0.0000
5	12.3E−04	2.36E−04	7.67E−04	16.9E−04	5.2167	0.0000
6	17.5E−04	3.32E−04	11.0E−04	24.0E−04	5.2639	0.0000
7	23.5E−04	4.46E−04	14.8E−04	32.2E−04	5.2688	0.0000

HIP_7: HIP data with follow-up of 7 years.

Table C-27: Estimated ACE(t) (95% CI) using the IGN-W model for the HIP_{18} data.

Method:	IGN-W		Data:	HIP_{18}		
Year	$\widehat{ACE}_{ign-w}$	SE_{cace}	LCL_{cace}	UCL_{cace}	Zvalue	$p-value$
1	0.32E−04	0.07E−04	0.17E−04	4.60E−05	4.2747	0.0000
2	1.11E−04	0.24E−04	0.65E−04	1.58E−04	4.6716	0.0000
3	2.33E−04	0.48E−04	1.40E−04	3.26E−04	4.8979	0.0000
4	3.92E−04	0.78E−04	2.40E−04	5.45E−04	5.0488	0.0000
5	5.88E−04	1.14E−04	3.65E−04	8.12E−04	5.1570	0.0000
6	8.19E−04	1.56E−04	5.12E−04	11.3E−04	5.2380	0.0000
7	10.8E−04	2.04E−04	6.82E−04	14.8E−04	5.3002	0.0000
8	13.8E−04	2.58E−04	8.74E−04	18.8E−04	5.3488	0.0000
9	17.1E−04	3.17E−04	10.9E−04	23.3E−04	5.3872	0.0000
10	20.6E−04	3.81E−04	13.2E−04	28.1E−04	5.4178	0.0000
11	24.5E−04	4.51E−04	15.7E−04	33.4E−04	5.4422	0.0000
12	28.7E−04	5.25E−04	18.4E−04	39.0E−04	5.4617	0.0000
13	33.1E−04	6.05E−04	21.3E−04	45.0E−04	5.4773	0.0000
14	37.8E−04	6.89E−04	24.3E−04	51.4E−04	5.4895	0.0000
15	42.8E−04	7.79E−04	27.6E−04	58.1E−04	5.4991	0.0000
16	48.1E−04	8.73E−04	31.0E−04	65.2E−04	5.5064	0.0000
17	53.6E−04	9.72E−04	34.5E−04	72.6E−04	5.5118	0.0000
18	59.3E−04	10.8E−04	38.2E−04	80.4E−04	5.5157	0.0000

HIP_{18}: HIP data with follow-up of 18 years.

Table C-28: Estimated survival values using the non-ignorable PPO survival models for the HIP_{18} data.

Data: HIP_{18}

Model:	NIGN-WW			NIGN-LL		
Years	$S_{c1}(t)$	$S_{c0}(t)$	$S_n(t)$	$S_{c1}(t)$	$S_{c0}(t)$	$S_n(t)$
1	0.9999	0.9999	0.9999	1.0000	1.0000	1.0000
2	0.9997	0.9997	0.9997	0.9998	0.9997	0.9998
3	0.9994	0.9993	0.9993	0.9995	0.9993	0.9994
4	0.9990	0.9988	0.9988	0.9991	0.9988	0.9989
5	0.9985	0.9981	0.9981	0.9986	0.9980	0.9982
6	0.9979	0.9974	0.9974	0.9979	0.9972	0.9974
7	0.9972	0.9965	0.9966	0.9972	0.9962	0.9965
8	0.9964	0.9956	0.9956	0.9963	0.9951	0.9955
9	0.9956	0.9945	0.9945	0.9954	0.9939	0.9943
10	0.9946	0.9934	0.9934	0.9944	0.9927	0.9932
11	0.9936	0.9921	0.9921	0.9934	0.9913	0.9919
12	0.9925	0.9908	0.9908	0.9922	0.9899	0.9906
13	0.9913	0.9893	0.9893	0.9911	0.9885	0.9892
14	0.9901	0.9878	0.9878	0.9899	0.9870	0.9878
15	0.9888	0.9861	0.9862	0.9886	0.9854	0.9863
16	0.9874	0.9844	0.9845	0.9873	0.9838	0.9848
17	0.9859	0.9826	0.9827	0.9860	0.9822	0.9832
18	0.9843	0.9807	0.9808	0.9846	0.9805	0.9817

Table C-29: Estimated survival values using the non-ignorable PPO survival models for the HIP_7 data.

Data: HIP_7

Model:	NIGN-WW			NIGN-LL		
Years	$S_{c1}(t)$	$S_{c0}(t)$	$S_n(t)$	$S_{c1}(t)$	$S_{c0}(t)$	$S_n(t)$
1	1.0000	0.9999	0.9999	1.0000	0.9999	1.0000
2	0.9998	0.9996	0.9998	0.9998	0.9996	0.9998
3	0.9996	0.9990	0.9994	0.9996	0.9990	0.9995
4	0.9993	0.9983	0.9990	0.9993	0.9982	0.9990
5	0.9989	0.9974	0.9985	0.9989	0.9973	0.9985
6	0.9984	0.9963	0.9979	0.9984	0.9963	0.9979
7	0.9979	0.9951	0.9972	0.9979	0.9951	0.9972

Table C-30: Estimated survival values using the F-R method for the HIP_{18} data.

Data:	HIP_{18}		
Model:		F-R	
Years	$S_{c1}(t)$	$S_{c0}(t)$	$S_n(t)$
1	0.9999	0.9998	0.9998
2	0.9997	0.9993	0.9996
3	0.9994	0.9989	0.9992
4	0.9991	0.9980	0.9985
5	0.9985	0.9967	0.9977
6	0.9979	0.9951	0.9969
7	0.9971	0.9945	0.9959
8	0.9960	0.9939	0.9949
9	0.9950	0.9924	0.9941
10	0.9940	0.9918	0.9930
11	0.9932	0.9907	0.9916
12	0.9922	0.9900	0.9896
13	0.9910	0.9888	0.9879
14	0.9898	0.9883	0.9860
15	0.9887	0.9870	0.9841
16	0.9872	0.9857	0.9825
17	0.9857	0.9840	0.9816
18	0.9845	0.9813	0.9809

Table C-31: Estimated survival values using the F-R method for the HIP_7 data.

Data:	HIP_{18}		
Model:		F-R	
Years	$S_{c1}(t)$	$S_{c0}(t)$	$S_n(t)$
1	0.9999	0.9998	0.9998
2	0.9997	0.9993	0.9996
3	0.9994	0.9989	0.9992
4	0.9991	0.9980	0.9985
5	0.9985	0.9967	0.9977
6	0.9979	0.9951	0.9972
7	0.9979	0.9951	0.9972

Table C-32: The ITT estimated survival values using the non-ignorable PPO survival models for the HIP_{18} data.

Data: HIP_{18}

Model:	NIGN-WW		NIGN-LL	
Years	$S_1(t)$	$S_0(t)$	$S_1(t)$	$S_0(t)$
1	0.9999	0.9999	1.0000	1.0000
2	0.9997	0.9997	0.9998	0.9998
3	0.9994	0.9993	0.9995	0.9994
4	0.9989	0.9988	0.9990	0.9988
5	0.9984	0.9981	0.9984	0.9981
6	0.9977	0.9974	0.9977	0.9972
7	0.9970	0.9965	0.9969	0.9963
8	0.9961	0.9956	0.9960	0.9952
9	0.9952	0.9945	0.9951	0.9941
10	0.9942	0.9934	0.9940	0.9928
11	0.9931	0.9921	0.9929	0.9915
12	0.9919	0.9908	0.9917	0.9902
13	0.9907	0.9893	0.9905	0.9887
14	0.9893	0.9878	0.9892	0.9872
15	0.9879	0.9862	0.9879	0.9857
16	0.9864	0.9844	0.9865	0.9841
17	0.9848	0.9826	0.9851	0.9825
18	0.9832	0.9808	0.9837	0.9809

Table C-33: The ITT estimated survival values using the non-ignorable PPO survival models for the HIP_7 data.

Data: HIP_7

Model:	NIGN-WW		NIGN-LL	
Years	$S_1(t)$	$S_0(t)$	$S_1(t)$	$S_0(t)$
1	1.0000	0.9999	1.0000	0.9999
2	0.9998	0.9996	0.9998	0.9996
3	0.9996	0.9992	0.9996	0.9991
4	0.9992	0.9986	0.9992	0.9985
5	0.9988	0.9978	0.9988	0.9977
6	0.9983	0.9969	0.9982	0.9968
7	0.9977	0.9958	0.9977	0.9958

Table C-34: The ITT estimated survival values using the F-R method for the HIP_{18} data.

Data:	HIP_{18}	Model: F-R
Years	$S_1(t)$	$S_0(t)$
1	0.9999	0.9998
2	0.9996	0.9994
3	0.9994	0.9990
4	0.9989	0.9982
5	0.9983	0.9970
6	0.9976	0.9957
7	0.9967	0.9950
8	0.9956	0.9942
9	0.9947	0.9930
10	0.9937	0.9922
11	0.9927	0.9910
12	0.9914	0.9899
13	0.9900	0.9885
14	0.9886	0.9875
15	0.9872	0.9861
16	0.9857	0.9846
17	0.9843	0.9832
18	0.9833	0.9812

Table C-35: The ITT estimated survival values using the F-R method for the HIP_7 data.

Data:	HIP_7	Model: F-R
Years	$S_1(t)$	$S_0(t)$
1	0.9999	0.9998
2	0.9996	0.9994
3	0.9994	0.9990
4	0.9989	0.9982
5	0.9983	0.9971
6	0.9977	0.9958
7	0.9977	0.9958

Table C-36: Estimated survival values using the IGN-W model for the HIP_{18} data.

Model:	IGN-W	Data:	HIP_{18}
Years	$S_{c1}(t)$	$S_{c0}(t)$	$S_n(t)$
1	0.9999	0.9999	1.0000
2	0.9997	0.9995	0.9998
3	0.9994	0.9990	0.9996
4	0.9990	0.9984	0.9994
5	0.9985	0.9976	0.9991
6	0.9979	0.9966	0.9987
7	0.9972	0.9955	0.9983
8	0.9964	0.9943	0.9979
9	0.9955	0.9929	0.9973
10	0.9946	0.9914	0.9968
11	0.9936	0.9898	0.9962
12	0.9925	0.9881	0.9955
13	0.9913	0.9862	0.9948
14	0.9901	0.9843	0.9941
15	0.9888	0.9822	0.9933
16	0.9874	0.9800	0.9924
17	0.9859	0.9777	0.9916
18	0.9844	0.9753	0.9907

Table C-37: Estimated survival values using the IGN-W model for the HIP_7 data.

Model:	IGN-W	Data:	HIP_7
Years	$S_{c1}(t)$	$S_{c0}(t)$	$S_n(t)$
1	1.0000	0.9999	1.0000
2	0.9998	0.9995	0.9999
3	0.9996	0.9989	0.9998
4	0.9993	0.9981	0.9997
5	0.9989	0.9970	0.9995
6	0.9984	0.9958	0.9993
7	0.9979	0.9943	0.9991

Table C-38: ITT estimated survival values using the IGN-W model for the HIP_{18} data.

Data:	HIP_{18}	Model: IGN-W
Years	$S_1(t)$	$S_0(t)$
1	0.9999	0.9999
2	0.9998	0.9996
3	0.9995	0.9992
4	0.9991	0.9987
5	0.9987	0.9981
6	0.9981	0.9973
7	0.9976	0.9964
8	0.9969	0.9955
9	0.9961	0.9944
10	0.9953	0.9932
11	0.9944	0.9919
12	0.9935	0.9905
13	0.9925	0.9891
14	0.9914	0.9875
15	0.9902	0.9859
16	0.9890	0.9841
17	0.9878	0.9823
18	0.9864	0.9804

Table C-39: ITT estimated survival values via the IGN-W model for the HIP_7 data.

Data:	HIP_7	Model: IGN-W
Years	$S_1(t)$	$S_0(t)$
1	1.0000	0.9999
2	0.9999	0.9996
3	0.9997	0.9992
4	0.9994	0.9986
5	0.9991	0.9979
6	0.9987	0.9970
7	0.9983	0.9959